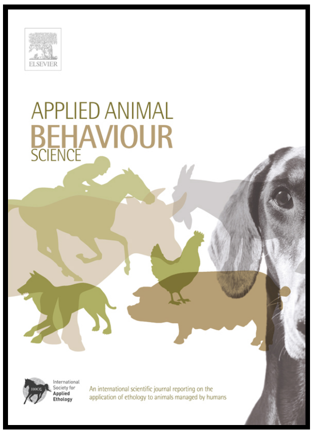
Journal Pre-proof



Influences of Indication Response Requirement and Target Prevalence on Dogs' Performance in a Scent-detection Task

Timothy L. Edwards, Claudia Giezen, Clare M. Browne

PII: S0168-1591(22)00115-0

DOI: https://doi.org/10.1016/j.applanim.2022.105657

Reference: APPLAN105657

To appear in: *Applied Animal Behaviour Science*

Received date: 15 February 2022
Revised date: 16 May 2022
Accepted date: 19 May 2022

Please cite this article as: Timothy L. Edwards, Claudia Giezen and Clare M. Browne, Influences of Indication Response Requirement and Target Prevalence on Dogs' Performance in a Scent-detection Task, *Applied Animal Behaviour Science,* (2022) doi:https://doi.org/10.1016/j.applanim.2022.105657

Influences of Indication Response Requirement and Target Prevalence

on Dogs' Performance in a Scent-detection Task

Timothy L. Edwards,[1] Claudia Giezen,[1] and Clare M. Browne[2]

[1]School of Psychology, University of Waikato

[2]School of Science, University of Waikato

Address correspondence to:

Timothy Edwards
School of Psychology
University of Waikato
Private Bag 3105
Hamilton 3240
New Zealand
Email: edwards@waikato.ac.nz

Abstract

Scent-detection dogs assist humans with many socially significant tasks and hold promise for assisting with many others. However, the methods used to train scent-detection dogs and the conditions under which they work are highly variable, and the influences of many relevant factors on scent-detection performance are poorly understood. Using an automated scent-detection apparatus that allowed the dogs to work independently from a handler, we evaluated the influence of two factors on scent-detection performance with amyl acetate as the target. In the first experiment in this study, we examined the influence of the indication response requirement on the performance of five dogs trained to perform a scent-detection task. The indication response consisted of the dogs breaking an infrared beam in a port through which they accessed samples. The response requirement was manipulated by adjusting the duration of the beam break that was required to activate food reinforcement if the target was present. As the indication response requirement increased, dogs' ability to detect the target remained unchanged, as indicated by stable log *d* measures across durations ($p = .09$), but their response bias, represented by log *B*, shifted from a tendency to indicate that targets were present to a tendency to indicate that targets were absent ($p < .001$). In the second experiment, we examined the influence of target prevalence, or the proportion of samples that are target samples, on four dogs' scent-detection performance. As target prevalence decreased, the dogs' ability to detect the target remained unchanged ($p = .13$), but their response bias shifted from a tendency to indicate that the targets were present to a tendency to indicate that targets were absent ($p < .001$). This finding aligns with the "low-prevalence effect," which is commonly observed in human signal detection research. The findings from both experiments have important theoretical and practical implications. For example, by adjusting response effort, we can alter dogs' response bias toward or away from indicating the presence of targets and, by having a clear understanding of the influence of target prevalence on bias, we can make informed decisions about the need to artificially increase target prevalence when dogs are searching for rare targets.

Keywords: canine, olfaction, response effort, sensitivity, specificity

1. Introduction

Domestic dogs (*Canis lupus familiaris*) are commonly trained to detect targets of interest in workplaces, airports, and other settings (Browne et al., 2006; Furton & Myers, 2001; Helton, 2009; Lorenzo et al., 2003). They have been trained to locate missing people and human remains (Hepper & Wells, 2005; Killam, 2004; Lorenzo et al.; Oesterhelweg et al., 2008), explosives (Furton & Myers; Gazit et al., 2005; Lazarowski & Dorman, 2014), narcotics (Jezierski et al., 2014; Lorenzo et al.; Nash, 2005), native and threatened animals (Browne et al., 2015; Cablk & Heaton, 2006), and pest species (Cooper et al., 2014; Gsell et al., 2009), for example. Dogs have also shown promise for detection of various cancers and other diseases (Edwards et al., 2017; Pirrone & Albertini, 2017).

The specific behavioral requirements associated with scent-detection tasks are important determinants of scent-detection performance. A typical approach to obtaining meaningful data from detection dogs involves training them to make a specific indication response when they encounter a target and to move on to the next potential target without "indicating" when they encounter a non-target.[1] This is achieved using discrimination training in which "hits" (indications of a target) are reinforced and "false alarms" (indications of a non-target) are not reinforced. If this training is successful, the target scent becomes a discriminative stimulus, which signals the availability of reinforcement for an indication response. As a result, the target evokes the indication response when encountered. Correct

[1]Some researchers have employed one-alternative forced-choice arrangements in which dogs are trained to identify a single positive sample in a group of samples but, as Edwards et al. (2017) discussed, it is not possible to obtain meaningful accuracy measures from this type of procedure for applied or translational research. Instead, each sample (or potential target) should be evaluated independently.

rejections (moving on without indicating when the target is absent) and misses (moving on without indicating when the target is present) are typically not followed by any specific consequence. Therefore, the signal detection payoff matrix associated with scent-detection work with dogs is generally as shown in Figure 1. This signal detection payoff matrix differs significantly from payoff matrices associated with standard yes/no discrimination procedures, which are commonly used in laboratory studies of factors influencing signal detection accuracy (e.g., Hume, 1974; Davison and Tustin, 1978). In these laboratory procedures, both hits and correct rejections are typically reinforced while misses and false indications are not reinforced.

<<Figure 1>>

Edwards et al. (2021) developed a laboratory model that more closely maps onto standard scent-detection protocols. Using domestic hens and a visual signal detection task, they evaluated the influence of three factors on signal detection performance in this task: the indication response requirement, signal prevalence, and the rate of reinforcement for hits. They found that indication response requirement was a critical determinant of bias, or the tendency to indicate or reject a potential signal, but that the other two factors did not play a critical role in determining signal detection accuracy with this procedure.

To clarify the applicability of these findings to the performance of scent-detection dogs, in this study, we evaluated the influence of the indication response requirement and signal prevalence on scent-detection performance in dogs in two experiments. Because manual training and testing procedures are susceptible to issues associated with cuing, data collection accuracy, and reliability and accuracy of reinforcement delivery (Edwards et al., 2017), in this study, we employed an automated approach, using a scent-detection apparatus and procedures as described by Edwards (2019). Following an initial training period, the dogs worked independently from the trainer by interacting with the apparatus directly. The apparatus automatically presented odor samples, reinforced correct indications according to a predetermined program, and recorded data.

When evaluating the influence of factors that may influence signal detection performance, it is useful to separate out changes in performance that are related to the detector's ability to detect the signal (i.e., stimulus bias) and the detector's tendency to indicate or not to indicate the presence of the signal when it is detected (i.e., response bias). Log *d* and log *B* have been used in similar studies to evaluate changes in stimulus bias and response bias, respectively (e.g., see Voss et al., 1993). These measures were used in the present analysis and are calculated as follows:

$\log d = 0.5*\log(H*CR/[M*FA])$;

$\log B = 0.5*\log(H*FA/[M*CR])$;

where H, M, FA, and CR are as defined in Figure 1.

2. Experiment 1: Indication Response Requirement

A key factor controlling bias (tendency to respond "yes" or "no") in the standard yes/no procedure is the payoff associated with responding "yes" or "no" in signal present and absent conditions (McCarthy & Davison, 1979). However, with models that are analogous to scent-detection models, there is no payoff associated with responding "no," so something else must prevent exclusive "yes" responding. Edwards et al. (2021) reasoned that the response requirement associated with the indication (i.e., "yes") response should influence bias. By manipulating the number of pecks required to count as an indication response and, therefore, to produce food when a target was present, they evaluated the influence of response requirement on signal detection accuracy and bias. The results demonstrated a clear relationship between the indication response requirement and bias, with hens producing higher hit rates and false alarm rates with lower indication response requirements and lower hit rates and lower false alarm rates with higher indication response requirements.

The specific behavioral function of the response requirement (i.e., response "effort") is not entirely clear, but Edwards et al. (2021) explored the following possibilities. In line with Fantino's (1969) Delay Reduction Theory, the time required to emit the response will produce a delay to the next reinforcer when the indication response is made and the signal is absent. By making a rejection response, this delay is reduced, and this reduction may serve as a reinforcer. A punishment interpretation of response effort does not appear to be theoretically sound because punishment must be delivered dependent on a response, and the response requirement is independent of the response (see Alling and Poling, 1995; Pinkston & Libman, 2017). Pinkston and Libman suggested that a behavioral economics approach to understanding the role of response effort may be fruitful. Response effort equates with cost in the behavioral economics framework, so predictions about interactions with this factor and others, such as demand elasticity, can be made. However, no comprehensive integration of behavioral economics and signal detection theory is yet available.

Essler et al. (2020) evaluated the scent-detection performance of dogs being trained to indicate the presence of blood plasma samples from individuals with ovarian cancer as a function of the dogs' type of indication response, a sit or stand-stare response. The "sit" dogs emitted 78 false indications while the "stand-stare" dogs emitted 48 false alerts. With only two dogs in each group, it was not possible to determine if this difference was statistically significant. There were significant differences between the two groups in the number of hesitations, defined as checking the sample twice or spending a long duration of time sniffing the sample without indicating; stand-stare dogs hesitate more often. Although the authors did discuss the role of the duration of the response requirement in their analysis, a response duration requirement was only imposed on one dog, which happened to be in the stand-stare group. Additionally, because the two different indication responses were qualitatively different, each associated with other differences in the dogs' experience in the task (e.g., stand-stare dogs had continuous access to sample characteristics while indicating), conclusions about the role of response effort in the obtained outcomes cannot be drawn from this study. A systematic analysis of the influence of this factor would be beneficial.

In the present experiment, using a within-subjects parametric design we evaluated the influence of the indication response requirement on scent-detection performance. An automated apparatus delivered food reinforcement only when a positive sample, containing amyl acetate, was present and the indication response requirement had been met. The response requirement, defined as breakage of an infrared beam spanning the sample port on the apparatus for a specified duration, was systematically increased to 13 s followed by re-exposure to a subset of previous duration values and two duration values that were lower than the initial duration value.

2.1. Experiment 1 Method

2.1.1. Subjects

Following preliminary screening for suitable temperament for laboratory work, 13 pet dogs were recruited initially. These dogs were assessed in introductory training sessions using the automated feeder that would be used in subsequent training and testing, and 5 dogs were selected for participation in the study based on their performance in these trials (see Table 1). None of the dogs had previous experience with the apparatus before participating in this study. To improve the reinforcing effectiveness of the dry food (kibble) that was dispensed by the automated feeder, owners were asked not to feed their dog within two hours of bringing them into the laboratory. Dogs were only housed intraday, typically for a morning or an afternoon period two days per week. Handling and care of the dogs was in accordance with an approved standard operating procedure. All experimental procedures in this study were approved by the University of Waikato Animal Ethics Committee (Protocol 1014).

Table 1

Dog Information

Name	Breed	Age (years)	Sex
Ruby	German shorthaired pointer x Labrador retriever	1	Female*
Trevor	German shorthaired pointer x Labrador retriever	0.5	Male
Katie	Blue heeler	9	Female*
Tui	Australian kelpie x collie x huntaway	1.5	Male*
Ella	Golden retriever	0.8	Female*

*Neutered

2.1.2. Apparatus

The dogs worked in a 3.2-m x 4.3-m room and interacted with a 1-m^3 apparatus containing a rotating carousel with 17 segments as described by Edwards (2019). Samples were placed inside each segment, and a lid covered the carousel to prevent cross-contamination between samples (Figure 2). The carousel rotated to present individual samples to the dog through a 10 cm sample port on the front side of the apparatus. To access each sample, dogs were required to place their nose in the sample port, breaking an infrared beam located on the inside of the front panel and push a flap on the front face of the aligned segment with their nose. The breaking of the infrared beam produced a continuous beep, which provided auditory feedback to the dog. The duration of the beam break was used to determine whether a sample was "observed" and "indicated" with observations and indications defined by an experimenter-specified value in the apparatus' control software.

<<Figure 2>>

Indications in the presence of the target activated a feeder, whereas indications in the presence of a non-target sample had no programmed consequences. An omnidirectional switch located on the right side of the front panel (Figure 2) became active once the dog had produced an observation response on any sample type (positive or negative). Once active, closing the switch turned the carousel, presenting the next sample to the dog. An automated feeder, a Treat and Train Remote Reward Dog Trainer™ manufactured by PetSafe, was located approximately 3.5 m away from the apparatus and dispensed dry kibble when triggered by the handheld remote control (e.g., during the shaping process) or by the apparatus when it was programmed to do so.

A computer in the adjacent room ran the custom software that controlled the apparatus. The status of the samples in the 17 chambers of the apparatus, the observation and indication response requirements, and other aspects of experimental sessions were entered into a configuration file that was read by the software. Using this information, the computer controlled rotations of the carousel and activation of the automated feeder. It also recorded all data coming from the apparatus associated with beam breaks and switch activations. Two cameras situated in the experimental room were used to stream footage of the sessions to a separate computer in the adjacent room that was used to monitor dogs and record the footage.

Between each session, the top and bottom plates of the apparatus were wiped with a 60% isopropanol solution. To prevent cross-contamination, the segments which contained positive samples were re-used only for positive samples in each new sequence until the segments were cleaned. At the end of each two-day training and testing period, all segments were washed with soap and water and wiped with the isopropanol solution.

2.1.3. Samples

All glassware was soaked in nitric acid, rinsed in de-ionized water, and dried in an oven prior to sample preparation. The positive sample solution was prepared with 0.25 mL of amyl acetate in 100 mL of deionized water. Negative sample material was prepared in the same way but without amyl acetate. Prior to experimental sessions, individual samples were prepared by adding 2 mL of the negative- or positive-sample solution to vials with a volume of 7 mL (6 cm tall and 1.5 cm wide). Adhesive labels were used to identify positive and negatives samples. Negative samples were prepared and placed into the apparatus first, according to a predetermined randomized sequence of 10 negative and 7 positive samples, followed by positive sample preparation and placement into the apparatus. All samples were placed in the apparatus and the lid placed on the carousel a minimum of five minutes before an experimental session began. Samples were used for no longer than two hours before being replaced with fresh samples.

2.1.4. Training

After the dogs learned to approach the feeder when it was activated remotely by the researcher, manual feeder activation was used to shape putting the nose into the sample port and open the segment flap for progressively longer durations. Prompts, such as pointing to the sample port, were used as necessary. Once dogs were reliably breaking the infrared beam, control of the feeder was transferred to the software. From this stage onward, training and experimental sessions were 17 trials long, one rotation of the 17-segment carousel. Dogs typically completed six sessions on each of the two days that they participated each week, with a short break (approximately three minutes) between sessions.

For initial scent training, all 17 segments were filled with 100-mL glass jars (instead of the 7 mL vials used for experimental sessions) filled with 4 mL of amyl acetate solution. The observation threshold was set at 200 ms (this was increased to 500 ms during experimental sessions) and the indication threshold required to activate the feeder was set at 500 ms. The indication threshold was gradually increased in 200-ms increments until a minimum positive sniff time of 2,000 ms was reached. Once the dog indicated on all positive samples over six sessions, limit switch activation was shaped using manual activation of the feeder. Once the dog activated the limit switch reliably, the carousel was loaded with 10 positive and 7 negative samples in alternating order. When a negative sample was present, an observation response to the sample, followed by a limit switch activation was required to progress to the next sample.

With prompting as necessary from the researcher, the dogs continued with this arrangement until they achieved 80% hit rate and correct rejection rate for 6 consecutive sessions. The researcher then gradually faded out all prompts, including removing themselves from the room. Novel randomized sequences of positive and negative samples were then introduced, with a ratio of 10:7, positive to negative. Once the dog reached 80% for both hit rate and correct rejection rate when working independently, the sample jar was changed from the 100 mL jar to the 7 mL vial. The amount of sample was reduced from 4 mL to 2 mL in increments of 0.5 mL. The concentration of the amyl acetate solution was systematically reduced (1:100 mL; 0.5:100 mL; 0.25:100 mL) to increase the difficulty of the task. The ratio of positive to negative samples was also systematically reduced from 10:7 to 7:10. The indication threshold was then adjusted incrementally until a threshold of 4,000 ms was reached.

2.1.5. Experimental Procedure

Starting with a baseline indication threshold of 4,000 ms to activate the feeder, the indication threshold was systematically increased by increments of 500 ms. If after six sessions, visual inspection of the data revealed a negative or positive trend, more sessions were completed until stability was achieved. If a dog took longer than two minutes to interact with the apparatus (i.e., break the infrared beam or close the limit switch), the session was terminated. If this occurred for three or more sessions consecutively, the increases of indication thresholds were terminated. Additionally, if a dog completed a session with hit rate below a value of 0.5 for six consecutive sessions, increases in indication thresholds were terminated. All dogs reached the 13,000-ms indication threshold without meeting termination criteria, and information about performance at higher threshold values did not appear to have any practical value. Therefore, after each dog reached 13,000 ms, the indication threshold was returned to two previously evaluated values, 6.5 and 9.5 s, to examine the possibility that the obtained data from the initial exposure were influenced by the systematically increasing order in which the thresholds were presented. Following this phase, dogs were exposed to two previously untested thresholds, 2 and 3.5 s, below the baseline threshold value (4 s).

2.1.6. Data Analysis

Hit rate (hits/[hits + misses]) and correct rejection rate (correct rejections/[correct rejections + false alarms]) were computed for each session. These values were used to make decisions about phase changes. Data from the last four sessions for each condition were used to summarize performance in each respective condition. These summary data were used to evaluate the effects of indication threshold on classification accuracy. Statistical analyses were completed using SPSS, Version 22.0, and R, Version 4.0.0.

2.2. Experiment 1 Results

Figure 3 displays the mean hit rate and correct rejection rate across all dogs as they were exposed to progressively increasing indication duration requirements. It also displays these rates during a second phase in which two previous duration values were revisited, and a final phase in which two duration values below the baseline value were employed. Hit rate remained high across all duration values, with a gradual decline as the duration requirement increased. Correct rejection rate was low (M = .67) at the initial 4-s duration requirement but increased rapidly as the duration requirement increased and stabilized around a high value from the 5.5-s duration requirement and above. An increase in correct rejection rate also corresponds with the increase from the 6.5-s to the 9.5-s duration requirement in the repeated exposures phase; correct rejection rate was low in the 3.5-s and 2-s duration requirement conditions in the final phase of the experiment but not as low as in the initial 4-s duration requirement condition at the start of the experiment. A logistic regression analysis based on a quasibinomial distribution was used to evaluate the influence of indication threshold on hit rate. Indication threshold was found to significantly predict hit rate (β = -0.11, t = -2.42, p

= .017). The same logistic regression analysis applied to correct rejection rate revealed a statistically significant influence of indication threshold on correct rejection rate (β = .17, *t* = 4.78, *p* < .001).

<<Figure 3>>

Log *d* and log *B* were plotted against the indication threshold values to evaluate the influence of the response requirement on stimulus bias and response bias, respectively (Figure 4). Higher log *d* values indicate higher stimulus bias (i.e., better discriminability); log *B* values above zero denote a tendency to indicate that a signal is present while values below zero indicate a tendency to indicate that a signal is absent. To produce log *d* and log *B* estimates when either false alarms (FA) or misses (M) in a condition were zero, which would produce an indeterminate outcome of the relevant calculations, the overall proportion of errors in the condition (FA/[H + FA + CR] or M/[H + M + CR]) was multiplied by the number of signal-absent (for H) or signal-present (for M) trials and the minimum of this number and 1 was used in place of the M or FA value. This same procedure was used by Voss et al. (1993) for this purpose. In conditions where both M and FA were zero, an estimate of .1 was entered for both M and FA values (this estimate is lower than any M or FA value otherwise obtainable). To evaluate the probability of obtaining the trends that we observed in log *d* (no apparent trend) and in log *B* (an apparent decreasing trend) across threshold values, we applied Kendall's trend test to each data set. As discussed by Elliffe and Elliffe (2019), this test makes no distributional assumptions, nor does it make assumptions about the scale of measurement. This test returned a non-significant result for log *d* across threshold values; ΣS = 103, *p* = .09.[2] The test returned a significant result for log *B* across threshold values; ΣS = =349, *p* < .001.

[2] Because of ties in the data (from the estimates produced when M and/or FA were zero), the test produced an undefined result; this result was rounded to the next defined value in the direction of zero (102 for the first test and -348 for the second test). This trend test was not applied to the hit rate and correct rejection rate data because there were too many ties (due to a ceiling effect).

<<Figure 4>>

The hit rates between the initial and the repeated exposure to 6.5-s and 9.5-s indication threshold values were compared to explore the potential influence of the initially systematic increasing threshold values. A Shapiro-Wilk test indicated the hit rate data were normally distributed. A paired-samples t-test was conducted to compare the hit rate from initial and repeated 6.5-s and 9.5-s indication threshold conditions. There was no significant difference between initial (M = .96, SD = .08) and repeated (M = .91, SD = .15) 6.5-s indication threshold hit rates, $t(4) = .703$, $p = .521$. There was also no significant difference between the initial (M = .85, SD = .16) and repeated (M = .85, SD = .24) 9.5-s indication threshold hit rates; $t(4) = .048$, $p = .964$). The correct rejection rates between the initial and the repeated exposure to 6.5-s and 9.5-s indication threshold values were also compared. A Shapiro-Wilk test indicated the correct rejection rates associated with the initial 9.5-s threshold condition were not normally distributed. A Wilcoxon signed rank test with continuity correction (to handle ties) was applied to the correct rejection rate data from the initial (M = .99, SD = .02) and repeated (M = .96, SD = .05) 9.5-s threshold conditions. The result was not statistically significant; $V = 7$, $p = 1$. A paired sample t-test was conducted to compare the two 6.5-s threshold conditions. There was not a significant difference in correct rejection rate between the initial (M = .95, SD = .05) and the repeated (M = .9, SD = .11) 6.5-s threshold conditions; $t(4) = 1.754$, $p = .154$.

2.3. Experiment 1 Discussion

Increasing the response requirement for an indication response had a systematic influence on the dogs' scent-detection performance. Hit rate declined and correct rejection rate improved as the indication threshold increased. The improvement in correct rejection rate was characterized by a rapid increase up to approximately the 6.5-s threshold value, followed by stable, high correct rejection rates. Hit rates declined gradually as the indication threshold increased. The absence of any trend in log *d* across the threshold values suggests that discriminability was not altered by the changing response requirement. The log *B* values shifted from a bias toward indicating the presence of the target (mean value > 0) at lower threshold values to a bias toward rejecting the target (mean value < 0) at higher threshold values, with log *B* values close to zero obtained in the 6.5-s to 7.5-s threshold range. These data support our hypothesis that response requirement should influence response bias but not stimulus bias, and align with the findings obtained by Edwards et al. (2021) in an analogous visual-detection procedure.

Analysis of the data from the repeated conditions suggested that the data obtained during the initial, systematically increasing threshold conditions were representative of data that would be obtained when those conditions were not presented in that specific order. The data from the final two conditions with threshold values below the initial (4-s) threshold value aligned with the general findings with respect to response bias. However, the correct rejection rate in these conditions was higher than the correct rejection rate in the initial, 4-s, threshold condition. We suspect that this may be a sequence effect as dogs in training typically start with a strong bias toward indicating samples as positive. We did not repeat the 4-s condition and, therefore, cannot confirm that this was the case in this experiment.

It appears that a ceiling effect may have attenuated the effect of the response requirement manipulation. We were unable to use a titration procedure in the present study to adjust the difficulty of the discrimination for individual dogs, as was done with individual hens by Edwards et al. (2019). For logistical reasons, all dogs were working with the same samples. If an airflow olfactometer were used instead (e.g., Hall & Wynne, 2019; Szyszka et al., 2012), sample concentration could be adjusted for individual dogs. Nevertheless, the predicted effect was obtained with an easily discriminable stimulus and, combined with the data from Edwards et al. (2019), we would predict that a more pronounced effect would be obtained in more difficult detection tasks.

3. Experiment 2: Target Prevalence

The prevalence of targets in a set of samples or in an environment where an animal is searching may influence their detection performance. To examine the influence of target prevalence on search behavior in dogs, Gazit, Goldblatt and Terkel (2005) repeatedly exposed explosives-detection dogs to two different search paths. The first path contained five targets in randomized locations, and a second path contained no targets. Dogs searched faster and required fewer verbal prompts on the first path compared to the second path. When a target was added to the second path on every fourth test day, search behavior did not improve, but when dogs searched a third path with a target also present on every fourth test day, search performance was better on this novel third path than on the second path. These findings suggest that extremely low target prevalence negatively influences search performance and that these prevalence effects are context-dependent.

Edwards, Ellis, et al. (2017) evaluated the influence of target prevalence on the detection performance of rats that were trained to identify sputum samples from tuberculosis-positive patients. In this study, the true status of “negative” samples was unknown; a subset of these samples was from individuals who were tuberculosis-positive but were not identified as such by the clinic where their sample was collected and evaluated by microscopy. Indications of these unknown-positive samples were not reinforced, but indications of known-positive samples were reinforced, so the rats were working under an intermittent schedule of reinforcement for hits. Edwards et al. systematically reduced known-positive sample prevalence and found that the rats’ hit rate began to suffer when known-positive prevalence reached 2%.

These findings align with the "low prevalence effect," a common effect obtained in signal detection studies with humans whereby hit rate declines as target prevalence is reduced (e.g., Wolfe et al., 2005). Target prevalence appears to impact response bias but not stimulus bias, with observers increasingly tending to indicate that a target is present as target prevalence increases (Wolfe & Van Wert, 2010). Although this effect is reliably obtained under the conditions that are commonly used to study it (i.e., visual search tasks with human subjects in laboratory settings) and in the aforementioned applied scent-detection studies, under other conditions, the effect has not been observed.

In Edwards et al.'s (2021) visual detection task with domestic hens (analogous to typical laboratory scent-detection tasks), after correcting for a systematic improvement in correct rejection rate as the experiment progressed, target prevalence was not found to influence stimulus or response bias. McCarthy and Davison (1979) evaluated the influence of target prevalence on stimulus and response bias in pigeons and found no significant influence of target prevalence on either measure.

Given that there is some evidence for a low-prevalence effect in previous scent-detection studies but poor evidence for this effect in a laboratory analogue to the general scent-detection procedures used for laboratory-based applied scent-detection with dogs, we aimed to determine if target prevalence influences scent-detection stimulus and response bias with this procedure by conducting a parametric evaluation of scent-detection performance across different target prevalence values.

3.1. Experiment 2: Methods

The same dogs used in Experiment 1 were used in this experiment, except Ruby, who was unavailable when this experiment was conducted. The same apparatus was used, and the same samples and method of preparation were used as described in Experiment 1. As in Experiment 1, dogs completed approximately six, 17-trial sessions per day, with approximately three minutes between sessions. The indication threshold was set at 5.5 s for all dogs. These procedures were approved under the same animal ethics protocol as described in Experiment 1.

3.2.1. Procedure

The positive to negative sample ratio was systematically manipulated. The same phase-change criteria used in Experiment 1 were applied in this experiment (i.e., six consecutive sessions with hit rate and correct rejections above 80% with no observable trend across the last four data points). Starting with a ratio of 7:10, the ratio of positive to negative samples was decreased systematically to 5:12, 3:14, and 1:16. The starting ratio of 7:10 was then re-presented, followed by systematically increasing proportions, 9:8, 11:6, 13:4, and 15:2. This was followed by a presentation of a selection of intermediate ratio values 4:13, 12:3, 2:15, and finally 6:11. This final sequence of ratios was examined to evaluate the possibility of order effects associated with the other sequences being presented in systematically increasing or decreasing order.

The termination criteria were the same as in Experiment 1; the session was terminated if a dog took longer than two minutes to interact with the apparatus. If this occurred for three or more sessions consecutively, sessions were stopped for the day. In addition to these criteria, if a dog had a hit rate below 0.5 for six consecutive sessions in a day, sessions stopped for that day. If this occurred for two days in a row, the dog's participation in the study ended.

3.2. Experiment 2: Results

Mean hit rate and correct rejection rate were plotted against the positive to negative sample ratio from lowest to highest (see Figure 5). The data from the second iteration of the repeated condition (7:10 ratio) were not included in this graph (a comparison of these values is provided below). Ratio values that were presented out of sequence, following the sequentially presented values, are indicated with an asterisk. No clear patterns are apparent from a visual analysis of the graph, including any patterns associated with the non-sequentially presented ratio conditions. Hit rate and correct rejection rate remained high across all ratio conditions. A logistic regression analysis based on a quasibinomial distribution was used to evaluate the influence of target prevalence on hit rate. Target prevalence was not found to significantly predict hit rate (β = -1.13, t = -1.28, p = .21). The same analysis applied to correct rejection rate also resulted in a statistically nonsignificant result (β = .14, t = .19, p = .31). Hit rates in the first (*M* = .97, *SD* = .02) and the second (*M* = .97, *SD* = .03) 7:10 ratio conditions were comparable, as were correct rejection rates in the first (*M* = .96, *SD* = .05) and second (*M* = .98, *SD* = .01) 7:10 ratio conditions.

<<Figure 5>>

Log *d* and log *B* were plotted against the ratio values to evaluate the influence of target prevalence on stimulus bias and response bias, respectively (Figure 6). The same procedure described in Experiment 1 was used to produce estimates of these values when either false alarms (FA) or misses (M) were zero. Except for a relatively low value in the lowest ratio condition (1:16) and a relatively high value in the 6:11 ratio condition (which happened to be the last condition to which the dogs were exposed), no orderly influence of ratio on stimulus bias (log *d*) is apparent based on a visual analysis of the graph. As in Experiment 1, Kendall's trend test was applied to the log *d* data; no statistically significant trend was detected with this test, ΣS = 34, p = .13. Log *B* values appear to be systematically related to the ratio values, with lower positive to negative sample ratios associated with a systematic bias toward indicating that the target was not present and higher ratio values associated with a systematic bias toward indicating that the target was present. Kendall's trend test indicated that this trend was statistically significant, ΣS = 146, $p < .001$.

<<Figure 6>>

3.3. Experiment 2: Discussion

In this experiment, we evaluated the influence of target prevalence on scent-detection performance in dogs. We found that target prevalence influenced response bias. With low target prevalence, the dogs displayed a bias toward indicating that the target was absent (log *B* values below zero), and with high target prevalence, the dogs displayed a bias toward indicating that the target was present (log *B* values above zero). The dogs' sensitivity to the target (log *d*) was not influenced by target prevalence. The low prevalence effect was not manifest in the hit rate or the correct rejection rate; these measures did not change as a function of target prevalence. As in Experiment 1, there appears to be a ceiling effect, with dogs performing with high accuracy in all conditions. This ceiling effect may have limited the range of these accuracy measures and, therefore, prevented the commonly observed decline in hit rate with lower target prevalence from occurring.

With their analogous visual discrimination task, Edwards et al. (2021) did not obtain clear evidence for a low prevalence effect, although they did obtain the highest log *B* values in conditions with the highest target prevalence (50%). The range of target prevalence values in that study (6% - 50%) was not as broad as the range presented in the present study (6% - 88%). In the present study, the most dramatic changes in response bias were observed across the highest three prevalence values. Edwards et al.'s analysis was also complicated by the hens' correct rejection rates improving gradually over the course of the study. In the present study, the dogs performed with the highest accuracy (log *d* and combined hit rate and correct rejection rate) in the final (6:11 ratio) condition to which they were exposed. This may be evidence that the dogs' accuracy was also improving as the study progressed. However, with the consistently high accuracy values obtained across all conditions, it is difficult to make this determination.

With the unbalanced payoff matrix that is associated with this and other scent-detection tasks (see Figure 1), one outcome of lower target prevalence is that there are fewer opportunities for reinforcement. Reinforcement only occurs when a target is present and an indication response occurs. Given that behavioral allocation tends to match available reinforcement (Poling et al., 2011), one possible outcome of an extremely low target prevalence is that a detector's behavior may be allocated elsewhere, and engagement with the detection task may deteriorate. None of the dogs in the present study met termination criteria, which suggests that the opportunities for reinforcement, even under the lowest target prevalence (6%) condition, were sufficient to sustain engagement with the scent-detection task. Edwards, Ellis et al. (2017) found that pouched rats' performance began to deteriorate when only 2% of samples represented opportunities for reinforcement. It is unclear how much further target prevalence could be reduced without impacting dogs' engagement in this scent-detection task; it is likely that this threshold would vary among individual dogs.

4. General Discussion

In this study, we evaluated the influence of the indication response requirement (response effort) and target prevalence on dogs' scent detection accuracy. In both cases, we found that these factors influence response bias but not stimulus bias. As the indication response requirement is increased, response bias shifts toward a tendency to indicate that the target is absent. As target prevalence increases, response bias shifts toward a tendency to indicate that the target is present. These findings have significant implications for scent-detection applications.

The response requirements of field-based scent-detection dogs are varied. For example, drug detection dogs have been trained to scratch or lie down, receiving reinforcement immediately upon handler confirmation of correct indications (Jezierski et al., 2014); explosives detection dogs have been trained to respond by sitting beside targets, with their handlers travelling a short distance to deliver reinforcement (Gazit et al., 2005); and bird carcass search dogs have been trained to stay, pointing and barking, until handlers arrive and provide reinforcement, as appropriate (Paula et al., 2011). Similarly, laboratory-based cancer detection dogs have been trained to respond to target presence by lying down, receiving human-delivered reinforcement for hits immediately (Rudnicka et al., 2014); and dogs employed to detect a specific human's odor can indicate by lying down, with reinforcement provided by handlers upon confirmation of correct responses (Marchal et al., 2016). The response effort in some of these examples (e.g., lying down while handlers rapidly assess the accuracy of the response) is low. Based on the findings from Experiment 1 of the current study, if false indications are problematic, increasing the indication response requirement (e.g., requiring a sit/stay for a longer duration) may help to improve this measure.

Target prevalence can be low in scent-detection applications, such as when searching for sparsely distributed species (e.g., critically endangered gorilla scats; Arandjelovic et al., 2015). Based on the findings from Experiment 2, under such conditions, one should anticipate a bias away from indicating the target is present. It is common practice in such circumstances for samples to be "planted" to artificially increase target prevalence, with the aim of providing opportunities for reinforcement and sustaining search performance (e.g., Wasser et al., 2004), but planting samples can be challenging as it is easy to introduce irrelevant scent cues in the process. Porritt et al. (2015) demonstrated that the practice of training dogs to indicate an additional target, which may be more practical to introduce to the operational environment, can facilitate scent-detection performance with the original target when it is rarely encountered in that environment. The present study provides additional clarification about the influence of target prevalence on specific scent-detection metrics, confirming that lower prevalence is likely to bias the detector away from indicating the presence of the target. These findings highlight the importance of having a clear understanding of the rate at which targets must be encountered to sustain accurate performance.

Scent-detection dogs frequently make crucial contributions to socially significant projects (Browne et al., 2006), but working with these dogs involves substantial time and financial commitments (Orkin et al., 2016). For these reasons, it is crucial to maximize detection dogs' effectiveness. The present findings may improve our ability to do so.

Acknowledgements: These experiments were conducted in partial fulfilment of Giezen's master's degree requirements. We gratefully acknowledge the technical support of Rob Bakker.

Funding: This work was supported by the University of Waikato.

References

Alling, K., & Poling, A. (1995). The effects of differing response-force requirements on fixed-ratio responding of rats. *Journal of the Experimental Analysis of Behavior, 63*(3), 331-346. https://doi.org/10.1901/jeab.1995.63-331

Arandjelovic, M., Bergl, R. A., Ikfuingei, R., Jameson, C., Parker, M., Vigilant, L. (2015). Detection dog efficacy for collecting faecal samples from the critically endangered Cross River gorilla (*Gorilla gorilla diehli*) for genetic censusing. *Royal Society Open Science, 2*, 140423. http://dx.doi.org/10.1098/rsos.140423

Browne, C., Stafford, K., Fordham, R. (2006). The use of scent-detection dogs. *Irish Veterinary Journal, 59*(2), 97-104.

Browne, C. M., Stafford, K. J., & Fordham, R. A. (2015). The detection and identification of tuatara and gecko scents by dogs. *Journal of Veterinary Behavior*, *10*(6), 496-503. https://doi.org/10.1016/j.jveb.2015.08.002

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

Cooper, R., Wang, C., & Singh, N. (2014). Accuracy of trained canines for detecting bed bugs (*Hemiptera: cimicidae*). *Journal of Economic Entomology, 107*(6), 2171-2181. https://doi.org/10.1603/EC14195.

Davison, M. C., & Tustin, R. D. (1978). The relation between the generalized matching law and signal-detection theory. *Journal of the Experimental Analysis of Behavior, 29*, 331-336. https://doi.org/10.1901/jeab.1978.29-331

Edwards, T. L. (2019). Automated canine scent-detection apparatus: Technical description and training outcomes. *Chemical Senses*, *44*(7), 449-455. https://doi.org/10.1093/chemse/bjz039

Edwards, T. L., Browne, C. M., Schoon, A., Cox, C., & Poling, A. (2017). Animal olfactory detection of human diseases: Guidelines and systematic review. *Journal of Veterinary Behavior, 20*, 59-73. https://doi.org/10.1016/j.jveb.2017.05.002

Edwards, T. L., Ellis, H., Watkins, E. E., Mulder, C., Mgode, G., Cox, C., & Poling, A. (2018). Tuberculosis detection by pouched rats: Opportunities for reinforcement under low-prevalence conditions. *Behavioural processes, 155*, 2-7. https://doi.org/10.1016/j.beproc.2017.05.015

Edwards, T. L., Tashkoff, A., Haycock, J., & Foster, T. M. (2021). A procedural analogue of prey detection and applied signal detection. *Behavioural Processes, 185*, 104356. https://doi.org/10.1016/j.beproc.2021.104356

Elliffe, D., & Elliffe, M. (2019). Rank-permutation tests for behavior analysis, and a test for trend allowing unequal data numbers for each subject. *Journal of the Experimental Analysis of Behavior*, *111*(2), 342-358. https://doi.org/10.1002/jeab.502

Essler, J. L., Wilson, C., Verta, A. C., Feuer, R., & Otto, C. M. (2020). Differences in the search behavior of cancer detection dogs trained to have either a sit or stand-stare final response. *Frontiers in Veterinary Science*, *7*, 118. https://doi.org/10.3389/fvets.2020.00118

Fantino, E. (1969). Choice and rate of reinforcement. *Journal of the Experimental Analysis of Behavior, 12*(5), 723-730. https://doi.org/10.1901/jeab.1969.12-723

Furton, K. G., & Myers, L. J. (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., Goldblatt, A., & Terkel, J. (2005). The role of context specificity in learning: The effects of training context on explosives detection in dogs. *Animal Cognition, 8*, 143–150. https://doi.org/10.1007/s10071-004-0236-9

Gsell, A., Innes, J., Monchy, P., & Brunton, D. (2009). The success of using trained dogs to locate sparse rodents in pest-free sanctuaries. *Wildlife Research, 27*(1), 39-46. https://doi.org/10.1071/WR09117

Hall, N. J., & Wynne, C. D. (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

Helton, W. S. (2009). *Canine ergonomics: The science of working dogs*. CRC Press. https://doi.org/10.1201/9781420079920

Hepper, P. G., & Wells, D. L. (2005). How many footsteps do dogs need to determine the direction of an odour trail? *Chemical Senses, 30*(4), 291-298, https://doi.org/10.1093/chemse/bji023

Hume, A. L. (1974). Optimal response biases and the slope of ROC curves as a function of signal intensity, signal probability, and relative payoff. *Perception and Psychophysics, 16,* 377-384.

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*, 112-8. https://doi.org/10.1016/j.forsciint.2014.01.013

Killam, E. W. (2004). *The detection of human remains* (2nd edition). Charles C Thomas Publisher.

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

Lorenzo, N., Wan, T. L., Harper, R. J., Hsu, Y.-L., Chow, M., Rose, S., & Furton, K. G. (2003). Laboratory and field experiments used to identify *Canis lupus* var. *familiaris* active odor signature chemicals from drugs, explosives, and humans. *Analytical and Bioanalytical Chemistry, 376*, 1212-1224. https://doi.org/10.1007/s00216-003-2018-7

Marchal, S., Bregeras, O., Puaux, D., Gervais, R., & Ferry, B. (2016). Rigorous training of dogs leads to high accuracy in human scent matching-to-sample performance. *PLoS ONE*, *11*(2), e0146963. https://doi.org/10.1371/journal.pone.0146963

McCarthy, D., & Davison, M. (1979). Signal probability, reinforcement and signal detection. *Journal of the Experimental Analysis of Behavior, 32*(3), 373-386. https://doi.org/10.1901/jeab.1979.32-373

Nash, M. (2005). Who let the dogs in? The use of drug sniffer dogs in mental health settings. *Journal of Psychiatric and Mental Health Nursing, 12*, 745-749. https://doi.org/10.1111/j.1365-2850.2005.00918.x

Oesterhelweg, L., Kröber, S., Rottmann, K., Willhöft, J., Braun, C., Thies, N., Püschel, K., Silkenath, J., & Gehl, A. (2008). Cadaver dogs – A study on detection of contaminated carpet squares. *Forensic Science International, 174*(1), 35-9. https://doi.org/10.1016/j.forsciint.2007.02.031

Orkin, J. D., Yang, Y., Yang, C., Yu, D. W., & Jiang, X. (2016). Cost-effective scat-detection dogs: Unleashing a powerful new tool for international mammalian conservation biology. *Scientific Reports, 6*, 34758. https://doi.org/10.1038/srep34758

Paula, J., Leal, M. C., Silva, M. J., Mascarenhas, R., Costa, H., & Mascarenhas, M. (2011). Dogs as a tool to improve bird-strike mortality estimates at wind farms. *Journal for Nature Conservation*, *19*(4), 202-208. https://doi.org/10.1016/j.jnc.2011.01.002

Poling, A., Edwards, T. L., Weeden, M., & Foster, T. M. (2011). The matching law. *The Psychological Record*, *61*(2), 313-322. https://doi.org/10.1007/BF03395762

Porritt, F., Shapiro, M., Waggoner, P., Mitchell, E., Thomson, T., Nicklin, S., & Kacelnik, A. (2015). Performance decline by search dogs in repetitive tasks, and mitigation strategies.

Applied Animal Behaviour Science, *166*, 112-122. http://dx.doi.org/10.1016/j.applanim.2015.02.013

Pinkston, J. W., & Libman, B. M. (2017). Aversive functions of response effort: Fact or artifact? *Journal of the Experimental Analysis of Behavior, 108*(1), 73-96. https://doi.org/10.1002/jeab.264.

Pirrone, F., & Albertini, M. (2017). Olfactory detection of cancer by trained sniffer dogs: A systematic review of the literature. *Journal of Veterinary Behaviour: Clinical Applications and Research, 19*, 105-117. https://doi.org//10.1016/j.jveb.2017.03.004

Rudnicka, J., Walczak, M., Kowalkowski, T., Jezierski, T., & Buszewski, B. (2014). Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography-mass spectrometry versus trained dogs. *Sensors and Actuators B: Chemical, 202*, 615-621. https://doi.org/10.1016/j.snb.2014.06.006

Szyszka, P., Stierle, J. S., Biergans, S., & Galizia, C. G. (2012). The speed of smell: Odor-object segregation within milliseconds. PloS one, 7(4), e36096. https://doi.org/10.1371/journal.pone.0036096

Voss, P., McCarthy, D., & Davison, M. (1993). Stimulus control and response bias in an analogue prey-detection procedure. *Journal of the Experimental Analysis of Behavior*, *60*(2), 387-413. https://doi.org/10.1901/jeab.1993.60-387

Wasser, S. K., Davenport, B., Ramage, E. R., Hunt, K. E., Parker, M., Clarke, C., & Stenhouse, G. (2004). Scat detection dogs in wildlife research and management: Application to grizzly and black bears in the Yellowhead Ecosystem, Alberta, Canada. *Canadian Journal of Zoology, 82*, 475-492. https://doi.org/10.1139/z04-020

Wolfe, J., Horowitz, T., & Kenner, N. (2005). Rare items often missed in visual searches. *Nature, 435*, 439-440. https://doi.org/10.1038/435439a

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*(2), 121-124. https://doi.org/10.1016/j.cub.2009.11.066

Figure Captions

	Indication Response	No Indication
Target Present	Hit (H) Reinforcement	Miss (M) No Reinforcement (Extinction)
Target Absent	False Alarm (FA) No Reinforcement (Extinction)	Correct Rejection (CR) No Reinforcement (Extinction)

Figure 1. *Payoff Matrix Associated with Scent-detection Tasks*

Figure 2. *Automated Scent-detection Apparatus.* Left: top view of the apparatus with the lid removed showing the 17 segments. Right: front panel with sample port in the middle and the omnidirectional switch on the right-hand side of the panel.

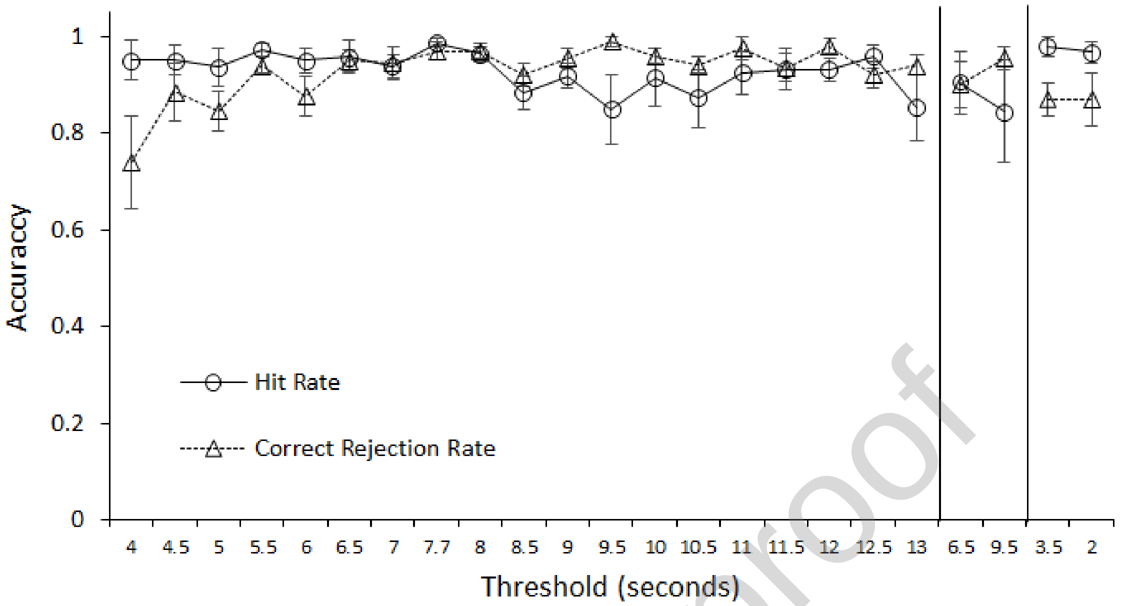


Figure 3. *Mean Hit Rate and Correct Rejection Rate at each Indication Threshold.* All dogs' (N = 5) mean hit rate (circles) and correct rejection rate (triangles) for each indication threshold, phase lines indicate when repeated threshold values and values lower than the initial threshold value were examined. Error bars indicate standard error of the mean.

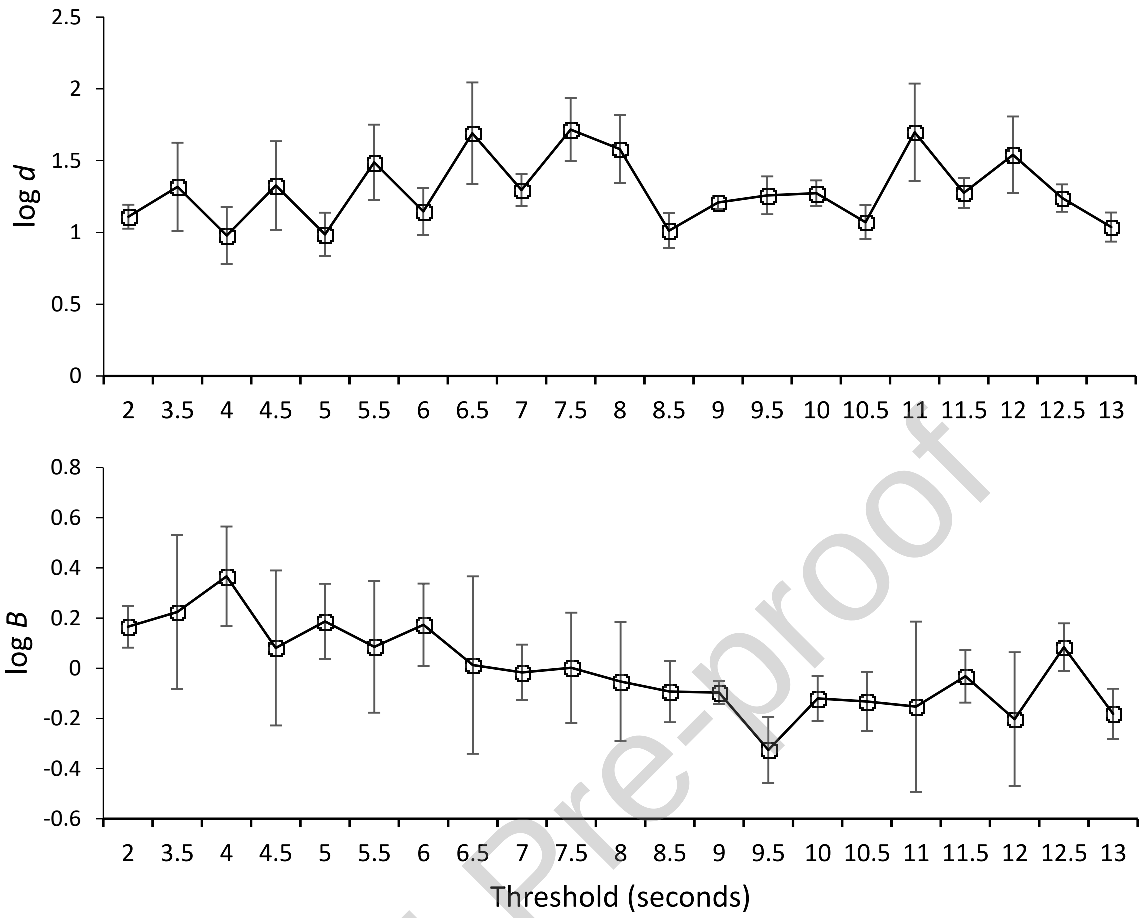
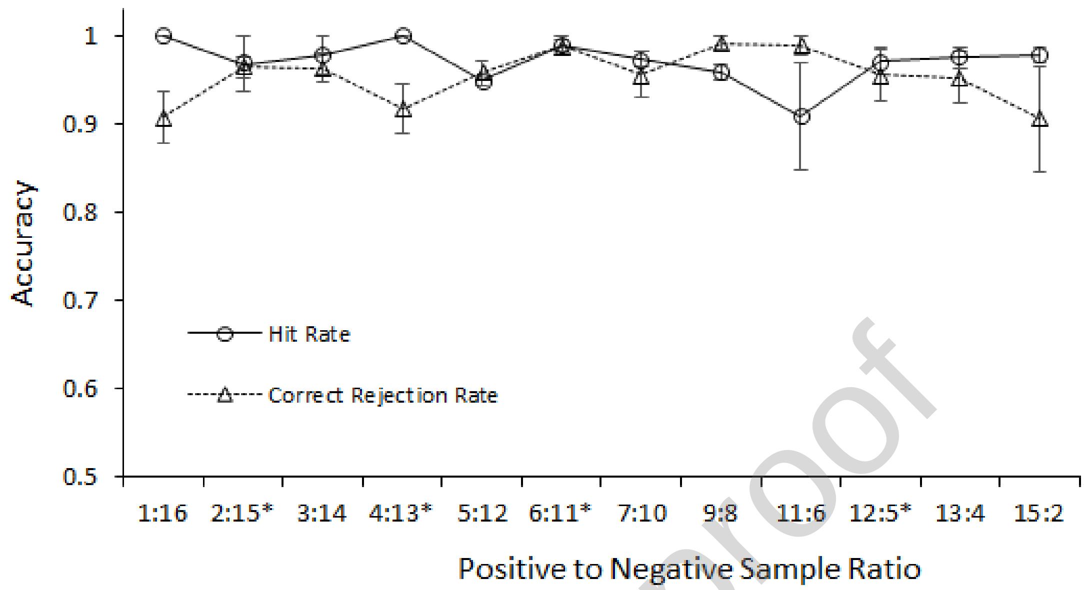


Figure 4. *Mean Log d and Log B Values at each Indication Threshold.* Log *d* values (top) and log *B* values (bottom) across each threshold value in ascending order. Error bars represent standard error of the mean.



¶

Figure 5. *Mean Hit Rate and Correct Rejection Rate Across Target Ratio Conditions.* All dogs' (N = 4) mean hit rate (circles) and correct rejection rate (triangles) for each target ratio condition in ascending order (not the order in which the conditions were presented; asterisks indicate those values presented out of order). Error bars indicate standard error of the mean. The y-axis starts at 0.5 to amplify the variability in the data.

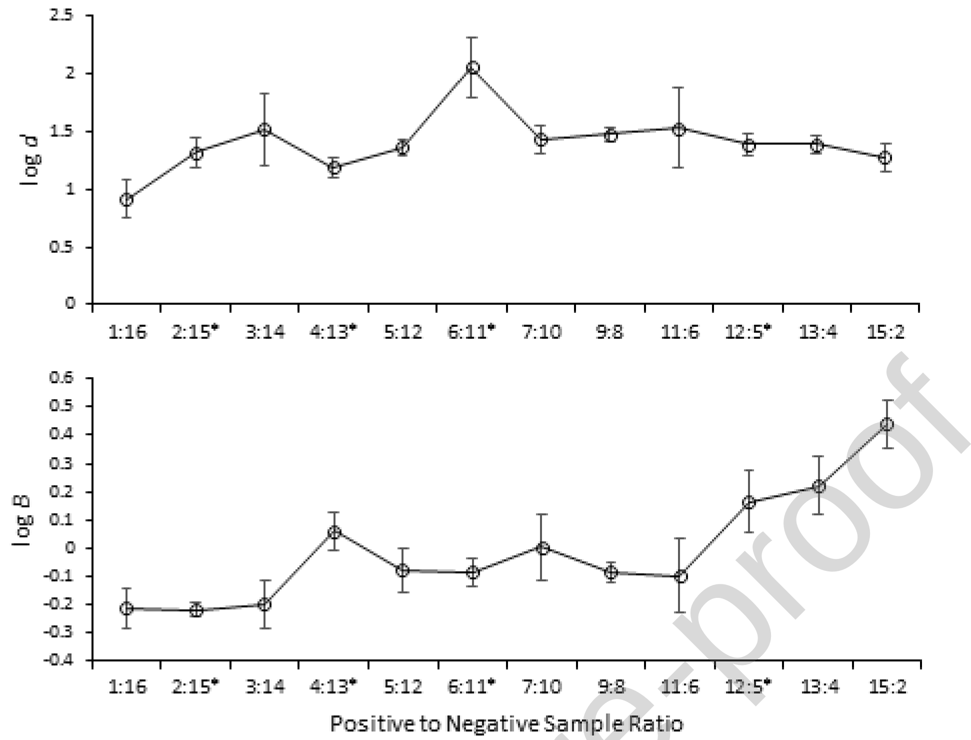


Figure 6. *Mean Log* d *and Log* B *Values Across Target Ratios.* Log *d* values (top) and log *B* values (bottom) across each ratio value in ascending order (not the order in which the conditions were presented; asterisks indicate those values presented out of order). Error bars represent standard error of the mean.

Highlights

- Dogs were tested with a range of scent-detection indication response requirements
- Higher response requirements biased the dogs away from indicating targets
- The dogs were also tested with a range of target sample proportions
- Lower sample proportions biased the dogs away from indicating targets
- Our findings suggest how these variables can be adjusted to maximize performance