http://researchcommons.waikato.ac.nz/

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author’s permission before publishing any material from the thesis.
The effect of response requirement and target probability on the performance of dogs during a go/no-go scent detection task

A thesis submitted in partial fulfilment of the requirements for the degree of Masters in Applied Psychology – Behaviour Analysis at The University of Waikato by CLAUDIA GIEZEN
Abstract

As the response threshold required to obtain a reinforcer is increased, the likelihood that the reinforcer will be obtained typically decreases. When a response is reinforced in the presence of a stimulus and not in its absence (i.e., a discriminated operant), the response threshold appears to influence the probability that the response will occur in the presence of the discriminative stimulus. In addition, when the target probability during a task is low, the number of responses performed typically decreases. We examined these two factors, response requirement and target prevalence, in a scent-detection task with domestic dogs where amyl acetate solutions were presented to each dog in an automated apparatus. The dogs were trained to place their nose in a sample port where they had access to samples, and to hold their nose in the port to indicate the sample was positive or to activate a limit switch when the sample was negative (a go/no-go scent detection task). During Experiment One, dogs were presented with various indication thresholds (seconds) which were manipulated systematically to evaluate the influence of response requirement on hit rate and correct rejection rate. During Experiment Two, target prevalence was manipulated systematically. Increasing indication threshold had a significant effect on hit rate when large increase in correct rejection rate as the response requirement was increased in the lower range and high correct rejection rate across most other values tested with a slight drop in hit rate as the response requirement was increased. Experiment Two revealed a small partial eta squared effect on accuracy when target prevalence was increased. Limitations and suggestions for future research are discussed.
Acknowledgements

Firstly I would like to thank Tim Edwards and Clare Browne for the opportunity to work in the dog lab and for supporting me throughout this whole process. Also for getting me out of my comfort zone and convincing me to present at the NZABA conference. To the dog owners who dropped and picked up their dogs, thank you for arranging your schedules to enable your dogs to participate, and for bearing with me when the study took an extra 8 months than originally planned.

A big thank you to Rob Bakker for fixing many bugs and glitches experienced with the scent detection apparatus and making it as easy to use. I would like to show my appreciation to the fellow scent detection students, especially Jesse and Margaret who would listen and de-stress me when needed, that was really helpful.

I would also like to thank the staff at student learning for helping me with my writing, the hardest part. My thesis would look very different if it weren’t for your help. I would like to share my gratitude to Simon, who dealt with many stressful moments when I lost references or sections, and for reading through the majority of my thesis to help with any grammar I may have missed. I would also like to acknowledge my sister, Erika, for influencing me to study psychology in the first place, a field I enjoy very much.
# Table of Contents

Abstract ........................................................................................................................................... i
Acknowledgements ....................................................................................................................... ii
Table of Contents ........................................................................................................................... iii
List of Figures ..................................................................................................................................... v
List of Tables ....................................................................................................................................... vii
Chapter 1 .......................................................................................................................................... 1
Introduction ......................................................................................................................................... 1
Chapter 2: Experiment One .............................................................................................................. 16
Introduction ......................................................................................................................................... 16
Method ............................................................................................................................................... 24
Subjects ............................................................................................................................................. 24
Apparatus .......................................................................................................................................... 25
Procedure .......................................................................................................................................... 28
Data analysis ...................................................................................................................................... 36
Results .............................................................................................................................................. 37
Discussion .......................................................................................................................................... 57
Chapter 3: Experiment Two .............................................................................................................. 64
Introduction ......................................................................................................................................... 64
Method ............................................................................................................................................... 70
Subjects ............................................................................................................................................. 70
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apparatus</td>
<td>70</td>
</tr>
<tr>
<td>Procedure</td>
<td>70</td>
</tr>
<tr>
<td>Data analysis</td>
<td>72</td>
</tr>
<tr>
<td>Results</td>
<td>72</td>
</tr>
<tr>
<td>Discussion</td>
<td>86</td>
</tr>
<tr>
<td>Chapter 4: General Discussion</td>
<td>91</td>
</tr>
<tr>
<td>References</td>
<td>96</td>
</tr>
<tr>
<td>Appendix A: Dog Initial Screening Form</td>
<td>102</td>
</tr>
<tr>
<td>Appendix B: Experiment One SOP</td>
<td>105</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1. The effect of signal strength on criterion thresholds.................................10

Figure 2. General matching law equation.................................................................12

Figure 3. Comparison between the rate of responding and rate of reinforcement in a GML equation.................................................................13

Figure 4. View of the custom-built apparatus used to present the samples to the dog.................................................................................................28

Figure 5. Trevor’s mean accuracy on Experiment One..............................................38

Figure 6. Tui’s mean accuracy on Experiment One.....................................................39

Figure 7. Katie’s mean accuracy on Experiment One................................................40

Figure 8. Ella’s mean accuracy on Experiment One....................................................41

Figure 9. Ruby’s mean accuracy on Experiment One................................................42

Figure 10. All dog’s mean accuracy on Experiment One..........................................43

Figure 11. Mean hit rate summary data for Experiment One.....................................45

Figure 12. Mean correct rejection summary data for Experiment One.....................46

Figure 13. ROC curve for all dogs.............................................................................47

Figure 14. ROC curve for individual dogs.................................................................48

Figure 15. Linear regression analysis of hit rate for all dogs...................................51

Figure 16. Linear regression analysis for correct rejection rate for all dogs.............54

Figure 17. Comparison of hit rate between ordered and unordered conditions....55

Figure 18. Comparison of correct rejection between ordered and unordered conditions...............................................................................................56

Figure 19. Trevor’s mean accuracy in Experiment Two.........................................73
Figure 20. Tui’s mean accuracy in Experiment Two..........................74
Figure 21. Katie’s mean accuracy in Experiment Two......................75
Figure 22. Ella’s mean accuracy in Experiment Two...........................76
Figure 23. All dogs mean accuracy in Experiment Two.......................77
Figure 24. Mean hit rate summary data for Experiment Two...............79
Figure 25. Mean correct rejection rate summary data for Experiment Two.....80
Figure 26. ROC curves for individual dogs.....................................81
Figure 27. ROC curve for all dogs collectively.................................82
Figure 28. Hit rate linear regression analysis of all dogs in Experiment Two.....84
Figure 28. Correct rejection rate linear regression analysis of all dogs in Experiment Two.........................................................86
List of Tables

Table 1. Response Matrix…………………………………………………………………………8

Table 2. Subject Information……………………………………………………………………..25

Table 3. Shaping Hierarchy………………………………………………………………………30

Table 4. Shapiro-Wilks Test for Normality Hit Rate Summary Data………………….50

Table 5. Shapiro-Wilks Test for Normality Correct Rejection Rate…………………..53

Table 6. Target Probability Examined…………………………………………………………...72

Table 7. Shapiro-Wilks Test for Normality Hit Rate Summary Data………………….83

Table 8. Shapiro-Wilks Test for Normality Correct Rejection Rate…………………..85
Chapter 1

Introduction

Trained animals help to detect various life-threatening targets, including explosives, diseases and narcotics (Helton, 2009; Weetjens et al., 2009). One organization that trains animals to complete such tasks is the Anti-Persoonsmijnen Ontmijnende Product Ontwikkeling (APOPO). APOPO is an organisation that has successfully trained giant African pouched rats (*Circetomys gambianus*) to detect various life-threatening targets, including landmines and tuberculosis. They have also evaluated the rats in human search and rescue scenarios (Edwards, Lalonde, Cox, Weetjens, & Poling, 2016; Lalonde et al., 2015), for detection of cigarette contraband in shipping containers (Mahoney, Lalonde, et al., 2014a), and for detection of salmonella in horse faeces (Mahoney, Edwards, et al., 2014). During training, pouched rats were trained using a “go/no-go” procedure, which is similar to most natural scenarios in that reinforcement follows correct positive indications but correct rejections are not followed by any programmed consequence (Weetjens et al., 2009).

Weetjens and his colleagues (2009) investigated whether giant African pouched rats were as accurate as traditional methods of tuberculosis (TB) detection, such as smear microscopy. Weetjens et al (2009) trained two rats to examine 10 randomised samples in a lineup, pausing for five seconds over TB-positive samples and ignoring TB-negative samples. Only indications in the presence of TB-positive samples were reinforced, and no consequences followed other response types. Both rats were able to discriminate between positive and negative samples with a hit rate of 73.1% and a correct rejection rate of 93%. The
authors concluded that applying group criteria (i.e., treating indications by either rat as a group indication) raised the hit rate (86.6%) but decreased the correct rejection rate (89.1%). Therefore a group of rats confirming the status of a sample may be more effective than a single rat’s confirmation. The researchers conducted the same experiment with an additional 16 rats and found that, over the 2,252 targets searched, 72% of the possible TB-positive targets met TB-positive group detection criteria (Weetjens et al., 2009). In addition, each rat had the ability to screen an average of 140 samples in 40 minutes as opposed to a microscopist screening an average of 40 samples a day. This finding suggests that animals may be a valuable addition to traditional screening for TB detection (Weetjens et al., 2009).

Poling and his colleagues (2010) trained giant African pouched rats to detect landmines, which are buried across 70 countries worldwide, resulting in an estimated 20,000 deaths per year (Edwards, Cox, Weetjens, T, & Poling, 2015; Mahoney, Lalonde, et al., 2014b; Poling et al., 2010; Verhagen, Weetjens, Cox, Weetjens, & Billet, 2006). Poling and his team (2010) trained 34 rats with the same procedure outlined earlier (Weetjens et al., 2009), with the indication response of pausing over a target for five seconds or more when presented with a pot filled with soil containing five drops of diluted trinitrotoluene (TNT; Poling et al., 2010). Once the response was acquired, the rat’s behaviour was shaped to detect five to seven mines in a 200 m² area. Once the rats were responding accurately, the rats were presented with the operational site which was a real testing field and thus contained an unknown number of targets. Therefore, because the status of the samples was unknown to the experimenters no
reinforcement was given during experimental sessions. Instead, to maintain performance, rats were exposed to simulated search environments with known targets to provide opportunities for reinforcement. The researchers found that out of the 93,400 m² searched, a total of 41 mines were located with no additional mines found with metal detectors. Although 617 false alarms were recorded, the number was no more than twice for a single rat on the same search site (Poling et al., 2010). These studies indicate that the use of animals alongside current methods can enhance detection of vital targets in real life scenarios (Poling et al., 2014).

Giant African pouched rats are common throughout Africa. However, they are not found naturally in other parts of the world. On the other hand, domestic dogs (*Canis lupus familiaris*) are found in most parts of the world and are commonly trained to detect particular targets found in real-world settings. For example, some studies have identified the dogs success in detecting targets in workplaces, airports, and forests (Browne, Stafford, & Fordham, 2006; Furton & Myers, 2001; Helton, 2009; Lorenzo et al., 2003), aiding police to locate people (Hepper & Wells, 2005), locating human bodies and remains (Killam, 2004; Lorenzo et al., 2003; Oesterhelweg et al., 2008), explosives (Furton & Myers, 2001; Gazit, Goldblatt, & Terkel, 2005; Lazarowski & Dorman, 2013), narcotics (Jezierski et al., 2014; Lorenzo et al., 2003; Nash, 2005), native and threatened animals (Browne, Stafford, & Fordham, 2015; Cablk & Heaton, 2006) and aiding conservation efforts including the eradication of pests (Cooper, Wang, & Singh, 2014; Gsell, Innes, Monchy, & Brunton, 2009; Smith et al., 2003). More recent applications of scent detection dogs have shown success in detecting various types of cancers (Cornu, Cancel-Tassin, Ondet, Girardet, & Cussenot, 2011; Edwards,
Helton (2009) reviewed the current literature on the use of dogs in scent detection tasks and concluded that detection dogs surpass the abilities of any known contemporary detection technology (Helton, 2009). For example, Pirrone and Albertini (2017) completed a study comparing the detection abilities of laboratory created E-noses and the olfactory abilities of dogs. The researchers found that E-noses could detect scents at the concentration threshold of 10 parts per billion, whereas the dog had the ability to achieve lower concentration thresholds between 10 to 6 parts per billion (Pirrone & Albertini, 2017).

Additional studies supporting the use of dogs as equal to or surpassing modern technology include Gialamas’ (1996) study on accelerant detection. Gialamas (1996) concluded that dogs were able to detect 0.1 µL of an accelerant on a burnt piece of carpet compared with traditional laboratory methods detecting between 0.1-0.5 µL of an accelerant on a given sample (Gialamas, 1996). It has also been discovered that dogs have the ability to not only locate a target but identify where a target had previously been (Oesterhelweg et al., 2008; Schoon & De Bruin, 1994), with the added ability to detect up to 10 different targets without a deterioration in performance (Williams & Johnston, 2002).

An additional study investigating the successful use of dogs was completed by Lazarowiski and Dorman (2013) who trained 16 dogs to differentiate between one component of an explosive by scratching, sitting, or lying by a target. The researchers then assessed if the dogs could discriminate the trained component when mixed with other chemicals.
Similar to the studies mentioned previously, correct responses in the presence of the target were followed by reinforcement, all other response types were ignored. The researchers provided a low target probability, by presenting one target only in a given session, to replicate conditions a dog would likely experience in a real-world environment. The researchers found that all dogs indicated a target was present 70-91% of the time. The study concluded that although the detection rate was not 100%, the probability of a target being detected still exceeded any current detection technology on the market (Helton, 2009; Jezierski et al., 2014; Lazarowski & Dorman, 2013).

The procedure used to train a dog is an important factor influencing the accuracy and performance of the animal as inaccuracy wastes time. In comparison with the go/no-go procedure, which provides natural search contingencies available in natural environments, the yes/no procedure provides reinforcement for both correct indication responses and also correct rejections (Voss, McCarthy, & Davison, 1993) The yes/no procedure was examined by Hume (1974) who, in an investigation of response bias, trained rats to push a right-hand lever (yes) when presented with a noise-plus-signal (target), and pushing the left-hand side lever (Cornu et al., 2011) when presented with noise alone. Both responses were reinforced when performed correctly, incorrect responses were punished with a time-out period. The strength of the noise in comparison with the signal was increased systematically to make the discrimination task more difficult. Hume (1974) found that when the signal became stronger, hit rate values increased along with a greater proportion of correct rejections. Alternatively, when the signal became weaker, the correct rejection rate increased and the number of incorrect responses (false alarms and misses; Table 1) increased (Hume, 1974).
However, the yes-no procedure has limitations as it does not replicate natural contingencies where correct rejections are typically not reinforced (Edwards, Browne, et al., 2017). Alternatively, the go/no-go procedure seems to be a better alternative for operational scent-detection work because the trainer only needs to know the status of some positive samples for training, whereas they would need to know the status of both positive and negative samples if a yes/no procedure was employed, which is more challenging logistically and more prone to error (Edwards, Browne, et al., 2017).

Voss and colleagues (1993) compared the accuracy obtained with a yes/no and a go/no-go procedure with pigeons. Pigeons were divided into two groups. The yes/no group were required to peck the left key in the presence of a bright illuminated light in the middle of a screen and to peck the right key when a dimmer light was illuminated. Both of the responses were reinforced with access to a feeder for three seconds. The second group of pigeons were presented with the same task, however, only a response to the left key was reinforced in the presence of the target, responding to the left key in the absence of the target and responding to the right key was not followed by reinforcement but was followed by a timeout period (all lights were turned off and responding to any key produced no consequence). The researchers concluded that the go/no-go procedure produced higher rates of accuracy with a greater bias towards the left key over the right key in comparison with the yes/no procedure. A second experiment was conducted which varied the timeout durations following incorrect responding and reduced the degree of brightness.
between the two illuminations, thus making it more difficult for the pigeons to
discriminate. The researchers found that in both procedures the rate of detection
dropped in the second experiment compared with the first experiment. However,
the go/no-go procedure produced a higher rate of accuracy compared than the
yes/no procedure. In addition, no significant changes resulted from increasing the
timeout duration on the right key in the go/no-go procedures (Voss et al.,
to the logistical benefits of the go/no-go procedure, it may also produce higher
accuracy (Voss et al., 1993).

Given this evidence, 100% accuracy is not always achieved, as false
alarms and misses can occur in applied settings. Therefore obtaining a better
understanding of scent detection can allow trainers and handlers to predict and
control the behaviour of a scent detection animal (Helton, 2009). Signal Detection
Theory (SDT) can be used to better understand how a responder makes a decision
by investigating the task used to differentiate and categorize unclarified stimuli
(noise) from a target (Abdi, 2007). A word find a puzzle, for example, requires an
individual to sort an array of letters (noise) to identify the target word “detection”.
Moreover, the theory provides a more detailed explanation of the decision-making
process, and how a particular subject or participant makes a decision (McNicol,
2005). As mentioned earlier, incorrect indications can occur in the presence or
absence of a target; therefore a total of four possible responses can occur, which
can be illustrated in a response matrix (Table 1). A hit refers to a positive
indication in the presence of a target stimulus, a miss involves a rejection response
in the presence of the target stimulus. In the absence of the target, a false alarm
refers to indicating a target is present in its absence, and a correct rejection involves a rejection response in the target’s absence (Abdi, 2007).

Table 1.

*Response matrix: Responses in the presence or absence of the target.*

<table>
<thead>
<tr>
<th>Response</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target present</td>
<td>Hit</td>
<td>Miss</td>
</tr>
<tr>
<td>Target absent</td>
<td>False alarm</td>
<td>Correct rejection</td>
</tr>
</tbody>
</table>

To evaluate the behaviour of a scent detection animal, SDT allows for the estimation of two main parameters. The first parameter measures an observer’s sensitivity ($d'$) to a target stimulus in comparison to noise (how much the target differs from the noise from the subject’s perspective) or the difficulty of a task (Abdi, 2007). For example, when the noise and the target are too close in value, the value of $d'$ is smaller. The second parameter ($\beta$) evaluates how biased an observer is in regard to the type of strategy they use to find a target (Abdi, 2007; Macmillian & Creelman, 1990). This can be explained by how likely one of the four possible responses from the response matrix is to occur. A liberal responder, for example, is more likely to say a target is present, resulting in more hits and a greater number of false alarms. Alternatively, a conservative responder is more inclined to say the target is not present, which will result in more correct rejections and misses, but fewer false alarms and hits (Blough, 2001).
While SDT measures an animal’s ability to differentiate between noise and a target, the true performance of a detector animal cannot be explained by using only one value. Instead, a combination of different measures of all responses made is required (Abdi, 2007). A liberal responder, for example, may appear to perform well with a high proportion of hit rates, however, they also produce a large proportion of false alarms. Alternatively, a subject may produce a low rate of false alarms, however, fail to recognise a large majority of targets (Helton, 2009). One measure used to understand the performance of a detector dog is hit rate, which is calculated by adding the number of hits divided by the sum of hits and misses (hit rate = hits/(hits + misses); Helton, 2009). The correct rejection rate is an additional measure which is calculated by dividing the number of correct rejections by the sum of false alarms and correct rejections (correct rejection rate = correct rejections/(false alarms + correct rejections); Helton, 2009).

In addition to accuracy measures, a “criterion threshold” can be used to predict how biased an observer is toward a particular response (i.e., liberal or conservative). This can help predict future responding by indicating how strong a target needs to be to increase responding to a particular response option, like a sensory threshold (Abdi, 2007). The criterion threshold can be plotted on a normal distribution curve (Figure 1). Any signal stronger than an observer’s criterion threshold will likely result in a positive indication response. Alternatively, a target weaker than the observer’s criterion will result in an observer ignoring that stimulus, resulting in a greater number of misses and correct rejections (Blough, 2001). For example, understanding a dog’s criterion threshold when detecting certain drugs will give handlers and trainers the ability to predict and control the behaviour of a scent detection animal in applied settings (Blough, 2001).
Another useful measure to predict and control the behaviour of future responding of a scent detection animal is response bias. Response bias is used to explain when there is a disproportional number of responses on one option. One reason for bias includes payoff which accounts for consequences attached to a response (Lynn & Barrett, 2014). For example, if only hits were reinforced, and no consequences followed a false alarm when an observer is uncertain of the presence of a target, an observer will most likely indicate a target is present as this results in a payoff compared to indicating a target is not present. Alternatively, if there is a punishment that follows a false alarm, an observer would more likely reject the target is present to avoid the punishment which follows a false alarm. An additional consideration in terms of response bias is the target probability within a scent detection task. Target probability can be used to account for the number of opportunities available for a reinforcement. If the target...
probability is low, a dog may reduce its searching behaviour altogether as the opportunity to gain a reinforcer is also low (Lynn & Barrett, 2014). An understanding of payoff and target probability is necessary to predict and raise performance levels.

In addition to SDT, the matching law can also add to the understanding of bias in responding. The matching law states that when one behaviour is reinforced more than an alternative, the probability of the response associated with a greater amount of reinforcement will also increase. Moreover, the amount of reinforcement increased corresponds to an increase in the probability of that response occurring (Reed & Kaplan, 2011). For example, if a response is reinforced with three pieces of food for pushing a lever on the right-hand-side of a screen and reinforced with one piece of food for pressing a lever on the left-hand-side, the animal will more likely push the right-hand lever. Additionally, the matching law can predict intervening schedules of reinforcement to decrease unwanted behaviour. For example, to reduce a dogs barking behaviour which is reinforced for every 6th bark, the matching law predicts that reinforcing an alternative behaviour on a richer schedule of reinforcement compared with barking behaviour, the dog will eventually respond with the alternative behaviour more often than barking (Reed & Kaplan, 2011).

To account for response bias in applied settings, Baum (1974) created the generalised matching equation (GML; Figure 2) to account for any influence on behaviour that is not a result of differing rates or schedules of reinforcement (Baum, 1974; Reed & Kaplan, 2011). In the GML equation, $B_1$ and $B_2$ represent the number of responses on either response alternative (e.g., left lever press and
right lever press). $R_1$ and $R_2$ represent the number of reinforcements delivered on each alternative, respectively. The $s$ represents the slope of the line of best fit, and $b$ represents bias, which accounts for any preference towards a certain response not accounted for by reinforcement alone (Baum, 1974; Reed & Kaplan, 2011).

$$\log\left(\frac{B_1}{B_2}\right) = s \log\left(\frac{R_1}{R_2}\right) + \log b$$

*Figure 2. General matching law equation*

When the sum of the GML equation (solid line in Figure 3) is compared to a perfect rate of responding with reinforcement rates being equal on both response options (broken line in Figure 3), an examination on how biased a responder can be measured (Figure 3). The top left graph in Figure 3 shows what would be expected if $b$ in the equation was equal to zero, meaning that the rate of reinforcement was equal on both response options. The top middle graph demonstrates what would occur if $b$ were greater than zero, indicating that there was an added bias toward the first response option that could not be explained by increased reinforcement on that response option. Alternatively, if $b$ is below zero, as shown in the top right of Figure 3, the responder would have a bias towards response option two that could not be explained by a greater rate of reinforcement for that response option (Reed & Kaplan, 2011).
Figure 3. Hypothetical comparison between the rate of responding and rate of reinforcement affecting bias during a scent detection task in accordance with a GML equation (Reed & Kaplan, 2011).

Referring back to the response matrix (Table 1), hits and correct rejections are responses which are reinforced in SDT, and so they can be adopted to stand as the two possible response options in the GML equation. Therefore, the probability of responding when presented with a target can be explained by the reinforcement ratio placed on either a hit or a correct rejection and not due to discrimination independent of the status of the target. The matching law offers a quantitative approach to understanding bias as a function of consequences associated with each quadrant in the SDT response matrix (Davison & Tustin, 1978).

It is acknowledged that the matching equation approach is clearly useful, but that, with go/no-go procedures, application of the equation results in the prediction that the animal will always respond go and rarely or never respond no-go because there is no reinforcement associated with such a response (Reed & Kaplan, 2011). Concha and colleagues (2014) suggested that when hits only are
reinforced, the dog will spend more time investigating positive samples compared with negative samples because correct responses in the presence of negative samples are not reinforced (Concha et al., 2014).

However, response options which are followed with no reinforcement (correct rejections, misses and false alarms) still occur. Because only hits are reinforced, reinforcement differences on responses cannot be used to explain the influence of correct rejections, misses and false alarms. For that reason, another factor may be influencing response bias during a go/no-go task. One unaccounted for factor may include the response required to produce either a go or no-go response, as the response requirement may effectively impose a penalty for false indications because no reinforcer is delivered to the indication response performed in the absence of the target (Weetjens et al., 2009). In Experiment One of the current study, the influence of response requirement on scent detection performance in dogs was investigated by increasing the indication threshold required to gain a reinforcer in the presence of the target.

Another factor impacting bias towards no-go responding may be target probability. The effect on go responding may decline as the target probability decreases because the target, which likely becomes as a conditioned reinforcer, is rare. In Experiment Two, the influence of target probability was examined by decreasing and increasing the proportion of positive and negative samples presented to dogs during a go/no-go scent detection task. The two factors of response requirement and target probability are relevant in nearly all scent detection applications, so
parametric analyses examining response characteristics as a function of these two factors would be useful.
Chapter 2: Experiment One

Introduction

Regardless of the type of signal detection task, an indication response is required, and every indication response is associated with some degree of effort (i.e., force required to push a lever, the duration of a nose hold, frequency of responses, or length of a distance required to gain a reinforcer). Zipf (1949) may provide a potential explanation for why response options that yield no reinforcement still occur with his derived principle of least effort. The principle of least effort states that all else being equal, an observer will choose a response that requires the least amount of physical effort over an alternative response which requires more physical effort (Zipf, 1949). For example, if two levers resulted in equal amounts of reinforcement, but lever one required two pushes and lever two required five pushes, an observer would most likely have a bias towards responding on lever one compared with lever two.

An investigation of the principle of least effort was completed by Chung (1965). During three experiments with four pigeons, Chung (1965) investigated the effects of increasing the force threshold required to peck a key in order to receive a reinforcer. During the first experiment, pigeons were presented with a response key with a baseline response threshold of 25 g which increased in increments of 25 g until 100 g was reached, then in increments of 50 g until a total 300 g of force threshold was achieved. The second experiment presented two keys which were both followed by 4 seconds of access to wheat. However, the force threshold differed randomly between the two keys. The last experiment kept the force
threshold between the two keys equal and both increased at an equal rate, so there was no difference between the keys. Overall, the researcher concluded that over the three experiments, increasing force threshold subsequently decreases responding to the response option associated with a greater force threshold (Chung, 1965).

To further investigate the effects of increasing response thresholds, Elsmore and Brownstein (1968) manipulated force thresholds on key pecking with three pigeons. The researchers varied the amount of force required to gain a reinforcer between 35 g and 175 g, and both keys were on an intermittent reinforcement schedule which provided a reinforcer approximately once every 2 minutes. The first pigeon had access to grain for 2.25 seconds when the yellow key was pecked and 4.50 seconds when the blue key was pecked. The second pigeon had access to grain for 2.25 seconds when the blue key was pecked and 4.50 seconds when the yellow key was pecked. The last pigeon had access to grain for 4.5 seconds when either coloured key was pecked. The number of pecks over 15 g of force and pecks that met the corresponding threshold was recorded for each pigeon. In support of the principle of least effort, the researchers found that a greater number of responses occurred during low force thresholds compared with high force thresholds. The researchers also observed that an equal number of responses on both response options above 15 g occurred, but the number of responses meeting the indication threshold decreased when 35 g or more was examined. This indicates that the number of responses on either key did not decrease, but rather the number of responses above the indication threshold decreased (Elsmore & Brownstein, 1968). The latter findings are also in correspondence with Pinkston and Libman (2017) who state that a higher
response threshold does not result in decreased responding to either response option. Rather, the number of responses reaching the criterion of the response threshold decreases (Elsmore & Brownstein, 1968; Pinkston & Libman, 2017).

An additional study examining the effects of increasing thresholds of force and number of pecks was conducted by Elsmore (1971). Pigeons learned to peck a red key with an associated reinforcement probability of 25% and a white key with an associated reinforcement probability of 50%. Two experiments were completed, the first manipulated the force threshold of pecks on both keys, and the second increased the number of pecks in a fixed-ratio schedule (Friman & Poling, 1995) threshold on both keys. The researcher concluded that when response thresholds were low, there was little difference in bias towards either key. However, when the response thresholds were increased, responding toward the red key, which had a smaller probability of reinforcement, resulted in decreased responding and little difference in responding to the white key was observed (Elsmore, 1971). In addition, Pinkston and Libman (2017) suggested that because there is no warning to the observer when the response threshold has increased, the increased response becomes aversive as previously reinforced responses are no longer reinforced. In turn, the new increment may create intermittent reinforcement because exceeding the previous threshold is not reinforced any longer (Pinkston & Libman, 2017).
Sumpter, Temple and Foster (1998) increased FR requirements to investigate whether larger FR requirements would result in a bias toward a shorter FR alternative. Six hens were trained to push a door with their head, then head, then to peck a key. The FR threshold then differed for both key and door response options independently. The FR schedule for key pecks was increased from FR 1 to FR 15 then to FR 50. The door responses varied from FR 1 to FR 3 then down to FR 2. A variable interval (VI) schedule was also manipulated randomly for both responses and differed between them (60 VI, 90 VI, and 180 VI) and was in place after achieving the examined FR requirement. The researchers found a greater number of key peck response compared to door pushes (four to one respectively) indicating that door pushing functioned as a higher threshold compared to key pecking. The results also indicated that when the FR threshold on key pecking was 5 times larger than the FR threshold on door pushes, hens will perform a greater number of door pushes over key pecks. A second experiment was conducted which placed a FR 1 threshold on both door and key responses but varied the force threshold required to gain a reinforcer to the door response option, with various VI schedules placed on both response options. The researchers determined that as the force threshold increased on door pushes, the number of door pushes decreased, but the different VI schedules saw a variation in these results (Sumpter, Temple, & Foster, 1998). The authors concluded that responses with different topographies can result in a bias toward one response option over another depending on the response threshold associated with the response option.

Alling and Poling (1995) increased the response threshold required to gain a reinforcer with rats where responding to only one response option was
reinforced. Responses on the left lever which met the response threshold were reinforced and no consequences followed responding on the right lever. Three experiments were conducted examining the effects of manipulating the FR and force required to push a lever. Experiment one involved a varied force condition where the amount of force required to depress a lever was increased systematically (25 g to 50 g, 100 g, and 200 g), and a constant force condition where the force threshold remained the same throughout the experiment (25 g). The researchers found that when the force threshold was examined at 50 g, 100 g and 200 g, overall responding decreased, whereas no effect was observed during the constant force condition. The second experiment investigated if the added force threshold of either 25 g or 200 g had an effect in relation to where in the position of the FR 15 requirement the force was added. The added force thresholds occurred during the 5th, 10th, or 15th response during the FR 15. The researchers concluded that where the increased response requirement occurred had no effect on responding. A final experiment examined whether different results would be observed during smaller FR schedules (FR 5 and FR 1). The same conditions as experiment one (varied force and constant force) were conducted in the final experiment. The researchers found that during the varied force condition, both FR schedules resulted in a decrease in responding and no difference was observed in the constant force condition, supporting the role of increased response requirement influencing responding (Alling and Poling, 1995).

Contrary to this analysis, some additional studies have also found counterintuitive results when introducing larger response requirements.
For example, Clement, Feltus Kaiser and Zentall (2000) trained pigeons to peck a white key on either a FR 1 schedule 50% of the time to gain access to a test pair discrimination which required a single peck, or an FR 20 on the remaining 50% of trials to gain access to the test pair. The test pairs presented a choice option between bright (target) and dim (non-target) illuminated keys. Meeting the corresponding FR requirement on either key resulted in reinforcement 50% of the time. The researchers discovered that when pigeons were given a choice between a FR 1 key option and a FR 20 option, the pigeon would choose the FR 20 option 62% of the time, indicating that the reinforcer following the larger FR 20 threshold on the centre key was more highly preferred over the same reinforcer followed by the FR 1 threshold key when given a choice. This may subsequently result in the harder to earn reinforcer being more effective than the easier to earn reinforcer (Clement, Feltus, Kaiser, & Zentall, 2000). Conversely, these findings contradict the principle of least effort, which states that a response with a smaller response threshold will produce more responding to that response option (Zipf, 1949).

Little research has been conducted to evaluate the effects of increasing response thresholds within a discrimination task using a go/no-go procedure. However, Tashkoff (2017) investigated the effects of response requirement during a go/no-go procedure by shaping hens to discriminate between two different light intensities. The brighter illumination was the target stimulus and the dimmer illumination was the negative stimulus. Following an observation response (a single peck on the lit key), a second, “advance” key was presented. At this point, the hen could continue to respond to the centre key or “advance” to the next trial by pecking the right key. A reinforcer was delivered once the FR threshold was
reached on the centre key when presented with the bright key. Meeting the FR requirement when presented with the negative stimulus and an incorrect response from the response matrix (Table 1) was followed by no reinforcement but a miss would result in the next trial being presented. The FR threshold was manipulated by increasing the number of pecks under an FR schedule on the centre key from a baseline of FR 10 which increased in increments of three until FR 22 was reached. Probe trials were also conducted to assess order effects and to determine where hit rate and correct rejection rate values were the highest. The researcher concluded that during baseline and smaller FR thresholds, hit rate was higher than correct rejection rate. However, this changed as the FR threshold increased and correct rejection rate values were greater than hit rate values, indicating that the hens were performing more misses and correct rejections compared with hits and false alarms as the FR threshold was larger. The results support the idea of a possible relationship between increases in response thresholds resulting in a decreases in responding to the corresponding response option, regardless of reinforcement rates.

The finding by Tashkoff (2017) also supports the idea that increased response thresholds can be used to understand response bias during a go/no-go task due to the delay to reinforcement, and during negative trials responding go would delay the presentation of the positive trial and therefore the opportunity for reinforcement. The researchers found a significant difference between the two response options when the response requirement on the go response option was set at FR 16, FR 19 and FR 22. Significant differences were also found between FR 22 and FR
10 and between FR 19 and FR 10. In addition, Tashkoff (2017) also concluded that differences in the order the FR requirements were presented to the hens had no effect on responding (Tashkoff, 2017).

A clear hypothesis can be formulated based on the work of Tashkoff (2017), and the studies mentioned throughout this chapter (Chung, 1965; Elsmore, 1971; Elsmore et al, 1968; Sumpter, et al, 1998; Zipf, 1949), but these hypotheses should be tested with dogs in a scent detection task. What is not yet clear is the influence of response thresholds on accuracy performance in dogs responding behaviour during a go/no-go scent detection task and few studies have investigated its influence. In the present study, dogs were trained to discriminate between amyl acetate and deionised water during a go/no-go procedure. The indication response of holding their nose in a sample port was initially set at 4 seconds and was systematically increased in increments of 0.5 seconds until 13 seconds was reached. Correct nose holds in the presence of the amyl acetate were reinforced with pieces of kibble. In the absence of the amyl acetate, dogs were required to activate a switch located next to the apparatus to proceed to the next sample. This response was not reinforced but was required to present the next discrimination to the dog. In accordance with the least effort principle and the results obtained by Tashkoff (2017), I hypothesised that the number of go responses would be high during baseline and smaller indication thresholds, but that go responses would decrease as the indication threshold increased and the number of no-go responses would increase.
Method

Subjects

All procedures in this thesis were approved by the University of Waikato’s Animal Ethics Committee (protocol number 1014). Thirteen pet dogs were recruited through Facebook and flyers around the University of Waikato campus and served as subjects for this study. Dogs were eligible to participate in the study if they were toilet trained, older than six months, and vaccinated against distemper, hepatitis and parvovirus. Appendix A displays the screening questions asked before a go was examined for their “trainability”.

Additionally, dogs were assessed for their “trainability” during an hour-long session where the dogs were trained to approach a feeder which delivered dry kibble when it was activated. All dogs that were unsuccessful in this task were not eligible to take part in the experiment. Each dog was held at the facility in a separate mesh dog crate or pen with access to water, a bed, and a toy. Five dogs were eligible to take part; Ruby, Trevor, Katie, Tui and Ella (Table 2).

All subjects had no previous experience with the apparatus before starting the experiment. To improve the reinforcing effectiveness of the dry food, owners were asked to withhold feeding their dogs their usual meal before coming. Owners gave written consent for their dogs to participate in the study after receiving information about the study from the researcher. Dogs were dropped off at the on-campus laboratory by their owners and were collected once sessions were finished.
Table 2.

*Subject Information*

<table>
<thead>
<tr>
<th>Name</th>
<th>Breed</th>
<th>Age (years)</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruby</td>
<td>German shorthaired pointer x Labrador retriever (crossbreed)</td>
<td>1</td>
<td>Female</td>
</tr>
<tr>
<td>Trevor</td>
<td>German shorthaired pointer x with Labrador retriever</td>
<td>0.5</td>
<td>Male</td>
</tr>
<tr>
<td>Katie</td>
<td>Blue heeler</td>
<td>9</td>
<td>Female</td>
</tr>
<tr>
<td>Tui</td>
<td>Australian kelpie x collie</td>
<td>1.5</td>
<td>Male</td>
</tr>
<tr>
<td>Ella</td>
<td>Golden retriever</td>
<td>0.8</td>
<td>Female</td>
</tr>
</tbody>
</table>

*Apparatus*

*Scent detection apparatus*

All experiments were conducted at a facility on the University of Waikato campus, in a 3.2 m x 4.3 m room where subjects were presented with a custom built 1-m³ apparatus containing a carousel with 17 segments. Samples were placed inside each segment individually, and a lid covered the carousel to prevent cross-contamination between samples (Figure 4a). The carousel rotated to present individual samples to the subject through a 10 cm sample port on the front side of the apparatus (Figure 4b). 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 (Figure 4b).
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,” which required a duration exceeding an experimenter-specified value and “indicated” when meeting or exceeding the set indication threshold.

Figure 4. Left, a, birds-eye view of the custom-built apparatus used to present the samples to the dog with the lid removed. One sample is placed in each of the 17 segments. Right, b, front panel with sample pot in the middle, and the omnidirectional switch on the right-hand side of the panel.

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 limit switch located on the right side of the front panel (Figure 4b) 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 to present the next sample to
the dog. The switch made a clicking sound to provide auditory feedback to the
dog when it had registered the response. No food reinforcement followed the
closing of the switch.

**Computer**

A computer in the adjacent room (Dell Optiplex 780 running on Windows Vista,
installed with original software developed for the apparatus) automatically
recorded data from the infrared beam circuit, including the time and duration of
the beam break, whether the sample in question was negative or positive, the total
length of the session, the number of reinforcers given, the amount of times the
limit switch was activated and whether each response was correct. If a go
response met threshold criteria in the presence of the amyl acetate, the feeder was
activated automatically by the computer programme, and no other response
activated the feeder.

**Feeder**

Located approximately 3.5 m away from the apparatus, the Treat and Train
Remote Reward Dog Trainer™ manufactured by PetSafe delivered dry food to the
dog when triggered by the computer or by a handheld remote control, which was
used during shaping. Chicken flavoured Pedigree adult dog biscuits were used as
the reinforcer. However, owners for two dogs (Ella and Trevor) brought in their
own dry food.

**Data collection and recording**

All sessions were video recorded with a Logitech 2 MP HD Webcam C600, which
streamed footage to a monitor outside the experimental room. This allowed the
researcher to witness and monitor all dog behaviour in the experimental room. Integrity checks were performed to ensure all procedures were followed as required.

**Procedure**

**Solution preparation**

All samples were prepared under a fume hood, and all glassware was acid washed prior to sample preparation (refer to *Cleaning*). Negative samples were created first to prevent cross-contamination between the negative samples of demineralised water and positive samples of amyl acetate. Latex gloves were used and changed between the creation of positive and negative samples and when needed after a spillage. Demineralised water was prepared in a 100 mL volumetric flask using a new plastic pipette and poured into a Schott bottle with Parafilm and a lid securing the sample. All negative samples were labelled “DM”.

Positive samples were prepared according to the same procedures as for negative sample preparation, with the addition of amyl acetate. The dilution used in experiment sessions was 0.025 mL in 100 mL of deionised water, however other dilutions were used during training (either, 1 mL, 0.5 mL, 0.12mL, or 0.06 mL of amyl acetate in 100 mL of deionised water). To prepare the positive samples, 0.25 mL of amyl acetate was added to a 100 mL volumetric flask. The whole flask was shaken slightly to ensure the sample was mixed well. The solution was added to a Schott glass bottle sealed with parafilm and a lid, and labelled “AA”. Solutions were stored for a maximum of two weeks before use. Bottles of the positive and
negative solutions were taken to the training facility where the solutions were used for sample preparation.

Sample preparation

Self-adhesive labels were placed horizontally on the outside of 7 mL sample vials (6 cm tall and 1.5 cm wide) to identify that those vials contained negative samples. To prepare the negative samples 4 mL of the negative solution (refer to Solution preparation above), was added to 10 of the labelled vials. All 10 negative samples were placed into a separate segment of the apparatus, according to a predetermined randomised sequence of 10 negative and 7 positive samples. Negative samples were placed in the apparatus before positive sample preparation began.

Positive samples were prepared following the same procedure; however, labels were placed vertically on the outside of the vials. The researcher placed 4 mL of the positive sample in the remaining seven vials, then placed each vial into the apparatus. All samples were placed in the apparatus for a minimum of five minutes before an experimental session began. No positive sample was left longer than two hours in the apparatus before the sample was disposed of and replaced with a fresh sample. If a sample spilt in the apparatus, any liquid was cleaned with propanol solution (80% propanol and 20% water) and a paper towel, and a clean vial containing a fresh sample would replace the spilt one.

Shaping

Shaping involved reinforcing successive approximations to the target behaviour of holding their nose in the sample port for 1 second when presented with a positive
target and activating the limit switch when presented with the negative target (Table 3).

Table 3

_**Shaping Hierarchy**_

<table>
<thead>
<tr>
<th>Level</th>
<th>Description of the step in the shaping hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Researcher delivers reinforcement with handheld remote control</td>
</tr>
<tr>
<td>2</td>
<td>Dog habituates to the sound of the feeder</td>
</tr>
<tr>
<td>3</td>
<td>Researcher uses prompts to encourage the dog to attend to the feeder (pointing and vocal encouragement)</td>
</tr>
<tr>
<td>4</td>
<td>Turns head away from the feeder</td>
</tr>
<tr>
<td>5</td>
<td>Turns body away from the feeder</td>
</tr>
<tr>
<td>6</td>
<td>Orientates toward the apparatus</td>
</tr>
<tr>
<td>7</td>
<td>Walks halfway across the room</td>
</tr>
<tr>
<td>8</td>
<td>Walks up to the front panel of the apparatus</td>
</tr>
<tr>
<td>9</td>
<td>Places nose anywhere on the panel</td>
</tr>
<tr>
<td>10</td>
<td>Places nose near the edge of the sample port</td>
</tr>
<tr>
<td>11</td>
<td>Places nose in the sample port</td>
</tr>
<tr>
<td>12</td>
<td>Breaks infrared beam</td>
</tr>
<tr>
<td>13</td>
<td>Places nose on the flap of the apparatus</td>
</tr>
<tr>
<td>14</td>
<td>All segments are filled with 17 positive samples with 4 mL of 1:1000 amyl acetate dilution in jars (5 cm tall and 4.5 cm wide)</td>
</tr>
</tbody>
</table>
Pushes the flap of the segment open slightly
Pushes segment flap halfway open
Pushes segment flap all the way open
Holds nose in the apparatus for 201-ms
Prompts are faded out
The duration of the nose hold is increased by 200-ms increments until 2001-ms is reached
Researcher moves to the right-hand-side of the apparatus, next to the omnidirectional limit switch
Reinforcement is delivered by the researcher
Prompts are used (vocal and pointing)
Head turning to the right
Front legs moving a step to the right
Both legs moving to the right
The dog's body orienting toward the limit switch
Placing any part of the body on the switch
Touching the switch
Pushing the switch
Moving the switch with enough force to make the switch click
Prompts are faded
Negative samples are placed interspersed in the apparatus with a positive:negative ratio of 10:7
Combining the two shaped responses
Researcher prompts nose in port on positive samples
Nose in port for 2001 ms in the presence of the positive sample is reinforced automatically by the apparatus. No consequences follow any response to negative samples.

Researcher prompts limit switch activation once the observation response is performed on negative samples.

Prompts are faded within at least 3 sessions.

Researcher moves towards the door a step at a time until by the door, within at least 3 sessions.

The door is open, researcher fades themselves out of the room, over at least 3 sessions.

The door is closed in small increments at a time, over at least 3 sessions.

The door is closed and the dog is performing the task alone.

A randomised sequence of the positive and negative samples are presented to the dog until 80% hit rate and correct rejection rate is achieved.

Novel randomised sequences of the positive and negative samples are presented to the dog.

Positive indication durations required to activate the feeder extended to 4,001 ms in 501 ms increments.

The sample is placed in a 10 mL vial from a 100 mL jar.

Amount of sample is reduced from 4 mL to 3 mL.

Amount of sample is reduced from 3 mL to 2 ml.

Dilution of amyl acetate is reduced from 1:1000 to 0.5:1000.
Dilution of amyl acetate is reduced from 0.5:1000 to 0.25:100

Proportions of positive samples are reduced from 10:7 to 7:10 (positive:negative)

The dogs were first desensitised to the feeder by the researcher activating the feeder with the handheld remote control while standing by the door of the experimental room. Once the dog was eating from and had habituated to the sound of the feeder, the researcher delivered reinforcement only when the dog moved further away from the feeder and towards the apparatus. The researcher gave prompts in the form of calling the dog’s name, pointing to the sample port and moving toward the front panel of the apparatus. The feeder was activated for a dog placing their nose in the sample port, breaking the infrared beam, and then for pushing the segment flap to gain access to the sample. Once the dog was reliably performing the observation response for 500-ms or more, all prompts were faded out.

During 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 (see sample preparation). When the apparatus was turned on lights illuminated along the border of the apparatus, these lights turned off at the end of the session. The observation threshold was set at 201 ms (this was increased to 501 ms during experimental sessions) and the positive indication threshold required to activate the feeder was set at 501 ms. The amount of time the dog needed to hold their nose in the sample port was gradually increased in
200 ms increments until a minimum positive sniff time of 2,001 ms was reached. Once the dog indicated correctly on 80% or more of the samples in a given session over six trials, the dog advanced to the next step in the shaping procedure, shaping of the limit switch activation response.

The researcher began the next step by standing next to the limit switch and reinforcing movements towards the limit switch by activating the feeder. Reinforcement was given for touching the limit switch at first, then for activating the limit switch (the specific topography was not selected). Once the dog performed the no-go response (i.e., activated the limit switch reliably 80% of the time for 6 consecutive sessions), the carousel was loaded with 10 positive and 7 negative samples in alternating order. The session was controlled by the computer from this point, but prompts were given as necessary by the researcher.

Once the dog was reliably performing the go and no-go responses correctly (i.e., 80% hit rate and correct rejection rate for 6 consecutive sessions), the researcher gradually faded out all prompts, including removing themselves from the room. The researcher took one step at a time to slowly remove themselves from the room. The door was closed in small increments until the door was completely closed and the dog was performing the task alone in the experimental room. Novel randomised sequences of positive and negative samples were then introduced, with a ratio of 10:7, positive:negative.

Once the dog reached 80% for both hit rate and correct rejection rate when working independently, the sample jar was changed from a 100
mL jar to a 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 (1:1,000 mL; 0.5:1,000 mL; 0.25:1,000 mL) to increase the difficulty of the task. Lastly, the ratio of positive and negative samples was systematically reduced from 10:7 to 7:10 positive:negative. Each step during the shaping of the smaller surface area and lower concentration dilutions, consisted of at least three consecutive sessions with 90% accuracy or more on both hits and correct rejections over six consecutive trials.

**Experimental sessions**

Starting with a baseline indication threshold of 4,001 ms to activate the feeder, the indication threshold was systematically increased by increments of 500-ms. Once the dog reached 80% or higher for both correct rejections and hit rate over six consecutive sessions, it was presented with a higher response requirement (i.e., current threshold plus 500 ms) to activate the feeder. If after six sessions and 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. If this occurred for two consecutive days, the dog was withdrawn from the study. 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. If this occurred for two days in a row, the dog was withdrawn from the study. 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 to examine possible order effects. Two reviewed indication thresholds of 6.5-s and 9.5-s and two unordered and untested values below baseline were examined to investigate shaping effects, refer to Appendix 1 for the standard operating procedure for this experiment.

**Cleaning**

At the end of each day, the apparatus was cleaned with a propanol solution to ensure there was no residual odour from the day’s sessions. All glassware used to prepare, store, and present the samples (e.g., sample jars/vials) were washed in an acid solution (NaOH) for a minimum of four hours, rinsed with deionised water, and placed in an oven set at 60°C for a minimum of four hours to dry. Windows were slightly open to allow ventilation into the room and food was securely contained, labelled and dated and all dogs were toileted regularly.

**Data analysis**

The mean hit rate and correct rejection rate values were computed for each indication threshold for each dog. These values were compared within and across all dogs. Accuracy obtained during the repeated exposure to specific indication threshold values was compared with accuracy obtained during the original exposure to the same values. ROC curves provide a summary of correct responding and incorrect responding to help gain a better understanding of an observer’s responding and is a common analysis completed in SDT to assess
responding accuracy (Maxion & Roberts, 2004). Therefore, ROC curves were created to assess characteristics of hit rate and correct rejection rate values across indication threshold values.

The hit rate for each dog on a given indication threshold was computed from the number of correct indications divided by the number of positive samples in a session (8 positive samples). The correct rejection rate was computed by the number of correct rejections a dog made in that session divided by the proportion of all negative samples presented (9 negative samples). Shapiro Wilks, Friedman’s, one-way repeated measures ANOVA, Wilcoxon ranked test and a Sign test were used to calculate normal distributions. Linear regression using the Durbin Watson statistic and a weighted least squares was used to calculate effect size and linear trend. Graphical calculations were performed using Excel 2013 for Windows and statistical analysis was completed on SPSS created by IBM Statistics for Windows, Version 22.0.

Results

All dogs completed the experiment (Figures 5, 6, 7, 8, 9 and 10). Once the indication threshold of 13 s was reached, no additional increases to the indication thresholds were evaluated; no dog reached termination criteria before 13 s indication threshold. Trevor’s correct rejection rate was lower than his hit rate at baseline, but, as the indication threshold increased, his correct rejection rate also increased until it was similar to his hit rate when a 7-s threshold was examined (Figure 5). The highest observed combined accuracy (hit rate = 1, correct rejection rate = 1) was obtained during the examination of the 11 s indication threshold. Similarly, high accuracy measures were obtained at 5.5 s and several others above
7.5 s. During the unordered thresholds, correct rejection rate decreased during 9.5 s and a slight increase in 6.5 s and hit rate increased to almost perfect accuracy, this was similar to the smaller than baseline thresholds examined.

Figure 5. Trevor’s mean hit rate (circles) and correct rejection rate (triangles) for each indication threshold, phase breaks show when unordered and smaller than baseline threshold were examined.

Tui’s hit rate and correct rejection rate were high during baseline and remained high when the indication threshold was set at 4 s to 6.5 s (Figure 6). Hit rate then decreased until the indication threshold was set at 10.5 s. From this point, hit rate increased to remain just below correct rejection rate during the examination of 11.5 and 13 s indication thresholds. Hit rate and correct rejection rate were the highest (1, 1) when 5.5 s and 6.5 s indication thresholds were examined. When unordered indication thresholds were reviewed, correct rejection rate values remained high, the values obtained at 6.5 s and 9.5 s were very close to the values obtained in the first exposure to these conditions. When 3.5 s and 2 s
indication thresholds were examined, Tui’s correct rejection rate was below his correct rejection rate, even more so for the 2 s indication threshold examined.

![Graph](image)

**Figure 6.** Tui’s mean hit rate (circles) and correct rejection rate (triangles) for each indication threshold, phase breaks show when unordered and smaller than baseline threshold were examined.

Similar to Trevor, Katie had a hit rate the same throughout, and the correct rejection rate started out low and increased as the threshold increases up to 5.5 s (Figure 7). Once the indication threshold surpassed 12 s, Katie’s correct rejection rate surpassed hit rate. During the reviewed durations, accuracy for both hit rate and correct rejection rate remained high. Hit rate remained high with a value of 1 at both reviewed 6.5 s and 9.5 s indication thresholds. Correct rejection rate values decreased in correspondence with shorter than baseline indication thresholds.
Figure 7. Katie’s mean hit rate (circles) and correct rejection rate (triangles) for each indication threshold, phase breaks to show when unordered and smaller than baseline threshold were examined.

During baseline and until the indication threshold was examined at 7.5 s, Ella’s correct rejection rate remained below hit rate (Figure 8). After the 7.5 s indication threshold, both hit rate and correct rejection rate remained between 0.85 and 1.0. When the indication threshold was set at 12 s, correct rejection rate was higher than hit rate. When previous indication thresholds were reviewed, correct rejection rate was just below hit rate at 6 s, then correct rejection rate increased to a value of 1.0 while hit rate decreased to a value of 0.88 during the 9.5 indication threshold examination. Once smaller than baseline indication thresholds were examined at 3.5 and 2 s, Ella’s correct rejection rate decreased, and her hit rate increased to a value of 1.0.
Figure 8. Ella’s mean hit rate (circles) and correct rejection rate (triangles) for each indication threshold, phase breaks to show when unordered and smaller than baseline threshold were examined.

During baseline, Ruby’s hit rate was at 1.0 and correct rejection rate was below 0.5, however, her correct rejection rate increased to 0.88 when the indication threshold was examined at 4 s (Figure 9). When the indication threshold was examined at 6 s, Ruby’s correct rejection rate surpassed hit rate; however, this was not consistent, as the correct rejection rate decreased during the examination of 6.5 s and 7 s indication thresholds. Hit rate and correct rejection rate were equal for indication thresholds 7.5 s and 8 s. Correct rejection rate fell below hit rate during the 8.5 s indication threshold, but then remained above hit rate for the majority of the higher indication thresholds. When previous indication thresholds were reviewed, correct rejection rate dropped to 0.65 for the 6.5 s and 0.5 for the 9.5 s indication threshold. When shorter than baseline indication thresholds were examined, hit rate increased to 0.93 and then 0.9 for the 2 s indication threshold.
Figure 9. Ruby’s mean hit rate (circles) and correct rejection rate (triangles) for each indication threshold, phase breaks show when unordered and smaller than baseline threshold were examined.

For all dogs, correct rejection rate was below hit rate for shorter indication thresholds between 4 s until 6.5 s (Figure 10). There was a near equal performance for both measures during the examination of the 7 s indication threshold. Correct rejection rate surpassed hit rate at 7.5 s indication threshold and near equal performance was observed for hit rate and correct rejection rate when 8 s and 8.5 s indication thresholds were examined. Correct rejection rate then surpassed hit rate values between 9 to 13 s indication thresholds. The highest combined value for correct rejection rate and hit rate was when the indication threshold was examined at 8 s with a combined value of 0.96. When previous indication thresholds were reviewed, near equal responding for both performance measures was observed for the 6.5 s threshold, and during the examination of the 9.5 s threshold the correct rejection rate was above hit rate value. Correct
rejection rate value then reduced below the hit rate value when the indication threshold was set at 3.5 s and 2 s.

**Figure 10.** All dog’s mean hit rate (circles) and correct rejection rate (triangles) for each indication threshold, phase breaks to show when unordered and smaller than baseline threshold were examined. Standard error bars are shown for each data point.

The number of sessions completed for each dog differed, as not all dogs reached advancement criteria within the same number of sessions. Therefore, summary measures for statistical analysis were computed using the data from the last four sessions in each phase for each dog. Figure 11 presents the average hit rates in the final four sessions for each dog at each threshold (Figure 11). Overall, the average hit rate values remained high during the experiment as values were never below 0.85. There appears to be a downward trend in hit rate during the 8, 9.5 and 13 s indication thresholds. Overall, individual dogs showed idiosyncratic differences across the indication thresholds. Katie, tended to perform above the mean hit rate for all dogs and Trevor, Tui, Ruby and Ella tended to perform below
the mean. The reviewed thresholds and the thresholds below baseline are consistent with the data obtained in the sequential conditions.

Figure 12 presents the overall mean correct rejection rate which was low during baseline and shorter indication thresholds (4.5, 5, 5.5, and 6.5 s) then increased and remained above a value of 0.9 for the remainder of the experiment (Figure 12). For individual dogs, correct rejection rate values showed idiosyncratic differences for each indication threshold. However, Katie and Trevor tended to perform above the mean. Ruby and Ella tended to perform below the mean and Tui tended to perform above the mean correct rejection rate until 11 s was examined where longer thresholds saw an increase in correct rejection rate. Visual inspection of the reviewed thresholds and the thresholds below baseline are consistent with the data obtained in the sequential conditions.
Figure 11. Mean hit rate summary data for each dog. The mean hit rate across all dogs for each indication threshold is also included (black line).
Figure 12. Mean correct rejection summary data for each dog. The mean hit rate across all dogs for each indication threshold is also included (black line).
Figure 13 displays the receiver operator characteristic (ROC) curve based on the mean hit rate and correct rejection rate across all dogs, and Figure 14 illustrates ROC curves for individual dogs. Hit rate is plotted against 1 – correct rejection rate for each of the original threshold values and the two values examined below baseline (2 s and 3.5 s).

Figure 13. The mean hit rate for all dogs plotted against 1 – correct rejection rate for each of the original threshold values. Four points have been labelled for reference.

For all dogs the data points are clustered in the upper left, which indicates that the dogs’ performance was far above chance performance (Figure 13).
Figure 14. Mean hit rate for individual dogs plotted against 1 – correct rejection rate for each of the original threshold values. Four points have been labelled for reference.
The summary data (last four scores from the raw data) was used to conduct a statistical analysis to test for differences in accuracy among threshold values. A test of normality using Shapiro-Wilk test indicated the indication threshold of 4.5 s, 7.5 s, 11.5 and 12.5 s were significant and therefore do not appear to be normally distributed (Table 4). A non-parametric Friedman test of differences among related measures was conducted and rendered a Chi-square value of $\chi^2(18) = 21.705$, $p = 0.245$ which was not significant ($p > 0.05$).

On visual inspection of the data (Figure 15), there appears to be a slight downward linear relationship between hit rate and increased indication threshold, data from indication thresholds below baseline were also included in this analysis and appeared to correspond with the overall trend (triangle markers on Figure 15). A simple linear regression was calculated to predict hit rate based on indication thresholds. A Durbin Watson statistic of 2.445 confirmed independence of the residuals. A visual analysis of the standardized residual scatterplot indicated that the assumption of homoscedasticity had not been violated. There was a medium effect of indication threshold on hit rate, $F(18,76) = 1.096$, $p = .373$, $\eta^2 = .206$. Indication threshold was not found to predict hit rate, and a positive slope was found ($\beta = .376$, $t = -1.025$, $p = .312$, 95% CI[.006, .019]). This yielded a predictive model for specificity performance ($S = 0.9884 -0.0069 \times$ Indication threshold). On visual inspection of the data (Figure 15), there appears to be a slight downward linear relationship between hit rate and increased indication threshold, data from indication thresholds below baseline were also included in this analysis and appeared to correspond with the overall trend (triangle markers on Figure 15).
Table 4

*Shapiro-Wilks Test of Normality for Hit Rate Summary Data*

<table>
<thead>
<tr>
<th>Indication threshold</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 s</td>
<td>0.034*</td>
</tr>
<tr>
<td>4.5 s</td>
<td>0.001*</td>
</tr>
<tr>
<td>5 s</td>
<td>0.123</td>
</tr>
<tr>
<td>5.5 s</td>
<td>0.683</td>
</tr>
<tr>
<td>6 s</td>
<td>0.850</td>
</tr>
<tr>
<td>6.5 s</td>
<td>0.051</td>
</tr>
<tr>
<td>7 s</td>
<td>0.406</td>
</tr>
<tr>
<td>7.5 s</td>
<td>0.001*</td>
</tr>
<tr>
<td>8 s</td>
<td>0.272</td>
</tr>
<tr>
<td>8.5 s</td>
<td>0.148</td>
</tr>
<tr>
<td>9 s</td>
<td>0.089</td>
</tr>
<tr>
<td>9.5 s</td>
<td>0.051</td>
</tr>
<tr>
<td>10 s</td>
<td>0.272</td>
</tr>
<tr>
<td>10.5 s</td>
<td>0.208</td>
</tr>
<tr>
<td>11 s</td>
<td>0.161</td>
</tr>
<tr>
<td>11.5 s</td>
<td>0.046*</td>
</tr>
<tr>
<td>12 s</td>
<td>0.615</td>
</tr>
<tr>
<td>12.5 s</td>
<td>0.024*</td>
</tr>
<tr>
<td>13 s</td>
<td>0.456</td>
</tr>
</tbody>
</table>

*Note.* Findings were considered significant at the $p<0.05$ level; significant findings are marked with an asterisk.
A simple linear regression was calculated to predict hit rate based on indication thresholds. A Durbin Watson statistic of 2.445 confirmed independence of the residuals. A visual analysis of the standardized residual scatterplot indicated that the assumption of homoscedasticity had not been violated. There was a medium effect of indication threshold on hit rate, $F(18,76) = 1.096, p = .373, \eta^2_p = .206$. Indication threshold was found to significantly predict correct rejection rate, and a negative slope was found ($\beta = .376$, $t = -1.025$, $p = .312$, 95% CI [.006, .019]). This yielded a predictive model for specificity performance ($S = 0.9884 - .0069 \times \text{Indication threshold}$).

![Figure 15. Linear regression analysis (broken line) with the values for individual dogs (open shapes) and the mean hit rate for all dogs (full shapes). The triangles represent thresholds examined below baseline.](image)

A test of normality using Shapiro-Wilk test indicated the indication threshold of 7.5, 9.5, 11, 1.5 and 12 seconds (Table 5) were significant and therefore do not appear to be normally distributed. A non-parametric Friedman's
test of differences among repeated measures was conducted and rendered a Chi-square value of \( x^2(18) = 41.84, p = .001 \) which was statistically significant (\( p < .05 \)). A Wilcoxon signed rank test showed that increasing the indication threshold did elicit a statistically significant change in correct rejection (\( Z = -8.463, p = .00 \)) A sign test was also conducted which indicated that increased indication threshold elicited a significant change in correct rejection rate (\( Z = -9.644, p = .00 \)).

On visual inspection of the regression line plotted through correct rejection data (Figure 16) there appears to be a positive relationship between correct rejection rate and indication threshold. A simple linear regression was calculated to predict correct rejection rate based on indication threshold. A Durbin Watson statistic of 1.815 indicated that there was independence within the residuals.

Based on a visual analysis of the standardized residuals scatter plot the assumption of homoscedasticity appeared to be violated. There was a large effect of indication threshold and correct rejection rate, \( F(18,76) = 2.867, p = .001, \eta^2_p = .404 \). Indication threshold was found to significantly predict correct rejection rate and a positive slope was found (\( \beta = -.211, t = -2.077, p = .04, 95\% \text{ CI}[-.211, -.284] \)). This yielded a predictive model for specificity performance (\( S = 0.8295 + 0.0117 \times \text{Indication threshold} \)).
Table 5

*Shapiro-Wilk Test of normality for Correct Rejection Rate Summary Data*

<table>
<thead>
<tr>
<th>Indication threshold</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 s</td>
<td>0.489</td>
</tr>
<tr>
<td>4.5 s</td>
<td>0.292</td>
</tr>
<tr>
<td>5 s</td>
<td>0.215</td>
</tr>
<tr>
<td>5.5 s</td>
<td>0.488</td>
</tr>
<tr>
<td>6 s</td>
<td>0.8</td>
</tr>
<tr>
<td>6.5 s</td>
<td>0.167</td>
</tr>
<tr>
<td>7 s</td>
<td>0.167</td>
</tr>
<tr>
<td>7.5 s</td>
<td>0.041*</td>
</tr>
<tr>
<td>8 s</td>
<td>0.542</td>
</tr>
<tr>
<td>8.5 s</td>
<td>0.412</td>
</tr>
<tr>
<td>9 s</td>
<td>0.506</td>
</tr>
<tr>
<td>9.5 s</td>
<td>0.001*</td>
</tr>
<tr>
<td>10 s</td>
<td>0.364</td>
</tr>
<tr>
<td>10.5 s</td>
<td>0.488</td>
</tr>
<tr>
<td>11 s</td>
<td>0.001*</td>
</tr>
<tr>
<td>11.5 s</td>
<td>0.028*</td>
</tr>
<tr>
<td>12 s</td>
<td>0.028*</td>
</tr>
<tr>
<td>12.5 s</td>
<td>0.363</td>
</tr>
<tr>
<td>13 s</td>
<td>0.695</td>
</tr>
</tbody>
</table>

*Note.* Findings were considered significant at the $p<0.05$ level; significant findings are marked with an asterisk.
Figure 16. Linear regression analysis (broken line) with the values for individual dogs (open shapes) and the mean correct rejection rate for all dogs (full shapes). Triangles represent thresholds examined below baseline.

A comparison of hit rate between the ordered and unordered indication thresholds of 6.5 and 9.5 s was analysed to examine shaping effects (Figure 17). The solid lines represent the mean hit rate for each indication threshold and the “R” after the indication threshold represents the unordered indication thresholds. Mean hit rate for 6.5 s decreased from 0.958 during ordered 6.5 s to 0.905 at the unordered 6.5 s indication threshold conditions. For 9.5 s, there was a downward trend between the ordered and unordered conditions with Ruby presenting as a potential outlier.
A test of normality using Shapiro-Wilk test indicated the indication threshold were normally distributed. A paired-samples t-test was conducted to compare the ordered and reviewed 6.5 s and 9.5 s indication thresholds. There was no significant difference between ordered ($M = .958, SD = .07588$) and the reviewed ($M = .905, SD = .14504$) 6.5 s indication threshold; $t(4) = .703, p = .521$. There was also no significant difference between the ordered ($M = .8495, SD = .15828$) and reviewed ($M = .8445, SD = .23622$) 9.5 s indication thresholds; $t(4) = .048, p = .964$.

*Figure 17.* Comparison of hit rate between 6.5 and 9.5 ordered and unordered indication threshold conditions.
A comparison of correct rejection rate between the ordered and unordered indication thresholds of 6.5 and 9.5 s was examined to assess shaping effects (Figure 18). The means between the ordered and unordered thresholds remained the same, showing no difference in correct rejection rate and no obvious differences were observed for individual dogs.

![Comparison of correct rejection rate between the 6.5 and 9.5 ordered and unordered indication thresholds](image)

**Figure 18.** Comparison of correct rejection rate between the 6.5 and 9.5 ordered and unordered indication thresholds conditions

A test of normality using Shapiro-Wilk test indicated the indication threshold of the 9.5 s ordered condition was significant and therefore does not appear to be normally distributed normality. A non-parametric Friedman's test of differences among repeated measures was conducted and rendered a Chi-square value of $X^2(3) = .556, p = .907$ which was not statistically significant ($p > .05$). A paired sample t-test was conducted to compare the ordered and unordered indication thresholds of 6.5 and 9.5 s. There was not a significant difference between the ordered 6.5 s threshold ($M = .95, SD = .05$) and the reviewed unordered 6.5 s indication threshold ($M = .9, SD = .105$) conditions; $t(4) = 1.754, p =$
There was also no significant difference between the ordered ($M = .99$, $SD = .022$) and reviewed ($M = .955$, $SD = .0542$) 9.5 s indication thresholds; $t(4) = 2.064$, $p = .108$.

**Discussion**

The aim of this experiment was to investigate the effects of increasing the duration requirement on hit rate and correct rejection rate during a go/no-go scent detection task. It was hypothesised that as the duration requirement increased, the number of go responses (hit rate) would decrease and the number of no-go (correct rejection rate) would increase. Five dogs were trained to perform the go/no-go scent detection task where a discrimination between amyl acetate (target) and deionised water (negative) was trained. The dogs’ indication response of placing their nose in a sample port was required in the presence of the target, this was reinforced, in the absence of the target the dogs were required to activate a limit switch in order to turn the apparatus to the next discrimination. The duration of nose holds in the presence of the target was systematically increased in 0.5 s increments starting with a baseline of 4 s until 13 s was reached.

Visual analysis of the relationship between hit rate and correct rejection rate suggests a decrease in hit rate as the indication threshold increases and an increase in correct rejection rate as the threshold increased (Figure 10). These results indicate that indication does have an effect on both performance measures. Overall, the results supported the hypothesis that increased indication duration requirements would decrease hit rates and increase correct rejection rates. However, for individual dogs, the indication threshold where correct rejection rate surpassed hit rate differed. For example, Tui’s correct rejection rate surpassed hit
rate when the 4.5 s threshold was examined, and Ella’s correct rejection rate surpassed hit rate during the 7.5 indication threshold (Figures 5 to 9). These differences could be explained on the basis that Tui is a very active dog and may find holding his nose in the sample port during the larger indication thresholds less reinforcing than investigating other things in his environment, or may find other things more reinforcing than the nose hold. This idiosyncratic difference may have influenced the early decrease in hit rate in comparison to the other dogs.

For individual dogs, Ruby and Tui performed in accordance with the hypothesis as the number of go responding was high while the indication threshold was smaller, but as the indication threshold increased, the number of go responses decreased and more no-go responses occurred. The point at which this difference occurred differed between the two dogs; however, Tui’s correct rejection rate surpassed hit rate once the 4.5 s threshold was examined (Figure 6) and Ruby’s correct rejection rate surpassed hit rate once the 9 s threshold was examined (Figure 9). Trevor, Katie and Ella’s results were also in accordance with the hypothesis as the correct rejection rate did rise from baseline and hit rate decreased (Figures 5, 7 and 8 respectively). However, the data for all dogs follows the pattern of the hypothesis (Figure 10), indicating that an overall trend was observed which was in line with the hypothesis.

The hit rate summary analysis indicates that hit rate values remained relatively consistent throughout the indication thresholds examined, with a slight decrease during 8.5 and longer durations (Figure
An effect was found between hit rate and indication threshold which indicates that indication did have an effect on indication threshold, and therefore should be a factor to consider for. Additionally, there was no significant difference between the ordered and the unordered conditions for hit rate, which indicates that ordered effects did not influence the data. Moreover a medium effect was found on hit rate due to indication threshold, which further supports the hypothesis on the influence of this factor during responding on the go/no-go scent task.

Looking at the correct rejection rate summary data for all dogs, correct rejection rate steadily increased as the indication threshold increased. Additionally, a large effect was found in the correct rejection rate due to indication threshold, with no significant difference between the ordered and unordered conditions examined. Thus the results suggest that the indication threshold strongly effected dog performance during a go/no-go scent task, and therefore future consideration from handlers/trainers when working with scent detection dogs.

Indication thresholds below baseline were also examined to further assess whether the same pattern of responding would be observed with lower indication thresholds. When the hit rate and correct rejection rate for all dogs for 2 s and 3.5 s indication thresholds were plotted with the regression analysis, the results from the smaller durations tended to be in line with the general results. Additionally, the smaller than baseline values were examined after the unordered indication thresholds, which tended to have correct rejection rate above hit rate values. When the smaller than baseline thresholds were examined, the values either decreased or increased in accordance with the hypothesis. This finding further supports the
findings are in line with the hypothesis and suggests that indication will have an effect on performance and can be used to predict the behaviour of a scent detection dog during a go/no-go scent detection task. During data collection, when a positive or linear trend was observed in the raw data, more sessions were completed to see either hit rate or correct rejections stabilise. Therefore, the number of sessions completed on an indication threshold differed across dogs.

These findings are in line with Chung (1965), Elsmore and Brownstein (1968), Elsmore (1971) and Sumpter, Foster and Temple (1998) who concluded that an increase in force threshold corresponded to a decrease in responding towards the increased force option (Chung, 1965; Elsmore & Brownstein, 1968; Sumpter et al., 1998). However, those studies had an influence of an equal contradicting reinforcement rate on the smaller threshold response option which may have influenced the results. Additionally, Chung’s (1965) results were not obtained from a discrimination procedure, so it was unclear if similar findings would be found in a discrimination task (Chung, 1965).

Alling and Poling (1995) on the other hand, had similar methods to the current study as reinforcement was only available on the increased response threshold option, and no consequences followed manipulations of the alternative response option. This indicates that a decreased responding on the response option associated with an increasing force and not due to reinforcement rates. However, the data were not obtained from a discrimination task the study did not provide a discrimination task so it
was unclear if similar findings would be found in a discrimination task (Alling & Poling, 1995).

The findings from this current study aligned with the findings from Tashkoff (2017) the overall hit rate was high and correct rejection rate was low during baseline and, as the response requirement increased, hit rate decreased and correct rejection rate increase until the two values switched and correct rejection rate values were above hit rate values. Additionally, Tashkoff (2017) used the go/no-go procedure where only hits were reinforced and correct rejections were not, but a no-go response was required to proceed to the next trial. The indication threshold increase was done sequentially and not randomised as in the present study, in addition, it appears that the procedure used by Tashkoff (2017) is a useful analogue of the scent detection procedure employed in the current study (Tashkoff, 2017). However, the findings from the present study are not in line with Clement, Feltus Kaiser and Zentall (2000), who found that pigeons would choose a larger response option over a smaller response option when given a choice. However, both response options provided in the Clement et al. (2000) study were reinforced, where the current study only reinforced the go option (Clement, Feltus, Kaiser, & Zentall, 2000).

Furthermore, the findings from the present study are also in correspondence with the principle of least effort (Zipf, 1949), because as the indication threshold increased for the go response option, the number of responses toward that option decreased. Therefore, in terms of the principle of least effort, as the response requirement increased, it became less probable that the go response was performed. Additionally, activating the limit switch was a less “effortful”
response option during the beginning of the experiment however no reinforcement was associated with it. As employed by Davison and Tustin (1978) a version of the matching law equation that incorporated the influences of response requirement might be of value as there was no competing rate of reinforcement on no-go response option which could be used to explain an increase in the number of no-go responses (Davison & Tustin, 1978).

No deterioration in performance was observed during the 13 s indication threshold and there were no practical reasons to continue to increase the response requirement. Therefore, different results may have been obtained in longer indication thresholds were observed. Furthermore, the current study adopted an ordered systematic increase of the indication thresholds. This may have caused order effects. For example, the majority of dogs’ hit rates decreased in the reviewed indication thresholds compared to the ordered thresholds which indicate that shaping effects may have influenced this decline. Yet, this appears to have been caused by a single dog whose performance declined during that period but this difference was not significant but should be considered in future research. Furthermore, randomizing the thresholds may have proven difficult for longer durations, and extinction effects may have been observed as a result but a systematic increase in threshold is perhaps the only way a higher threshold would be implemented in practice and therefore seemed like a logical approach to evaluating the influence of threshold.

Moreover, from the present data, it appears that an individual dog’s indication threshold should be individually assessed. However, the current
data suggests that there are indication thresholds which tend to produce high performance across all dogs, which could serve as default indication threshold until individual assessment is conducted or if not practical. Future possible research could be completed to examine applied targets such as chemicals from explosives, biological samples for disease detection or narcotics to see if similar results could be obtained. The reader should bear in mind that only two reviewed thresholds and two thresholds below baseline were examined. Therefore, it would be interesting for future research to review or revisit a larger number of previously examined indication thresholds to ensure no shaping effects influenced results. During the course of the experiment, the apparatus and SOP’s were new and under development and hence some interruptions occurred, which might have resulted in more variability in the data than would have been obtained under more stable conditions. Additional future research could examine influences of response requirement in a yes/no scent detection task and whether similar results would be obtained.
Chapter 3: Experiment Two

Introduction

When targets are rare, opportunities for indication responses to be reinforced are also rare. Target prevalence may influence the probability of indications in the presence of the target and correct rejections in the absence of the target. Therefore, understanding the effect of low and common target probability and the possible effects of this on scent detection performance is important for scent detection research and applications.

McCarthy and Davison (1979) evaluated the influence of target probability on accuracy with pigeons using the standard yes-no procedure. The pigeons were presented with either a bright or dim centre key which, when pecked, produced a green right key and a red left key. When the target bright centre key was illuminated the pigeons were required to peck the red key, and when presented with the dim light, a peck to the green key was required. The probability of the presentation of the bright illumination (target) was manipulated, along with the value of varied ratio (VR) schedules of reinforcement for responding on each key. During the first condition, a varied 10% to 90% target probability was arranged along with a VR 3 schedule of reinforcement which resulted in either 3 seconds of access to wheat from a magazine when the VR was met, or the magazine lit up without wheat delivery. Similar to the first experiment, the second experiment arranged 10% to 90% target probability across trials, with reinforcement conditions consistent with a VI schedule of 60 seconds for both response options. The final experiment had a constant 70% target
probability of either response option, but the VI schedule was varied between a VI 12, VI 30, VI 90 and VI 96 for each response option. The researchers discovered that any differences in responding were controlled by the variation in reinforcement, rather than by target probability. This conclusion was drawn because only Experiments 1 and 3, where reinforcement was manipulated, produced significant changes in detection accuracy. This was in comparison to Experiment 2 where target probability alone was manipulated, and resulted in no significant differences in responding. (McCarthy & Davison, 1979). This study provides some insight into the possible effects of target probability. However, because both yes and no responses were reinforced, different results may be obtained under the go/no-go procedure. Therefore, within the go/no-go procedure, the question of target probability still remains.

In the yes/no procedure, opportunities to gain reinforcement are not necessarily influenced by target probability, but in the go/no-go procedure the influence of target probability also corresponds to the opportunity to gain a reinforcer during a go/no-go task because only correct responses in the presence of the target are reinforced. For example, when the target probability is decreased, the number of reinforcers available will also decrease. Often, explosive detector dogs who work under go/no-go conditions, may encounter almost zero targets in real-world settings, which corresponds to a low reinforcer probability. Therefore, low target probability can be dangerous if it results in deteriorated performance measures, as misses can be life-threatening.

To investigate this issue, Gazit, Goldblatt and Terkel (2005) reinforced correctly sitting at a trained explosive target with five dogs for, and all other
response types were ignored. During the first experiment, all dogs were repeatedly exposed to two paths daily. The first path, Path A, contained five known targets in randomized locations, and a second path, Path B, containing no targets. The researchers found that Path A resulted in quicker searching behaviour compared to Path B, and more verbal prompts were needed in Path B. To assess if occasionally planting a target on Path B would lead to an increase in search behaviour, a second experiment was completed on Path B only. Path B then contained a target on every fourth day only. No increase in search behaviour was observed and the percentage of hits significantly decreased compared to the probability of hits on Path A during Experiment 1. These results suggest that exposure to one target in four days was not sufficient to maintain search and indication behaviour in familiar environments. To see if the results from the first two experiments were context-based, a third experiment was completed in a novel environment, Path C. Similar to the second experiment, Path C contained one target every fourth day which resulted in a greater number of hits with an increase in search behaviour compared with Path B in Experiment 2. (Gazit et al., 2005).

The results from the above study (Gazit et al., 2005) suggest that a drop in hit rate and search behaviour during Experiment 2 was a result of previous experience on path B and not because of the present target probability (Gazit et al., 2005). However, the dogs had a lower hit rate on Path A in Experiment 1, which contained 5 targets, compared with Path C in Experiment 3, which had only one target, indicating that a lower target probability increased hit rate.
The current literature on target probability suggests that best practice involves planting known targets in applied environments to maintain responding. However, when component chemicals of explosives are target scents, this practice can be dangerous. Therefore Porritt and colleagues (2015) investigated whether training dogs to detect non-explosives (i.e., Vanillin, potassium chlorate) in applied settings would alter the detection rates and accuracy of explosive targets. To examine this, 21 dogs were trained to detect one explosive target and one non-explosive target. Once this was achieved dogs were assigned to groups. The first group Zero-Target group were not presented with any targets in a work context, the second group One Target group were presented with only one non-explosive target, and finally, a third Three Target group were presented with three explosive targets (Porritt et al., 2015). First, the three groups were presented with their respected condition, then after 44 days, they were reintroduced to the test phase which included explosive targets, all groups were exposed to four explosive targets in a given session. The researchers then compared the performance between the three groups to assess the effectiveness of the different types of planted targets during the work phase. The results from the Zero-Target group were similar to the results from Path B during Experiment One in the Gazit et al. (2005) study, as search behaviour decreased and hit rate and search behaviour took longer to recover during test sessions where dogs were re-exposed to trained explosive targets (Porritt et al., 2015).

The researchers also concluded that the group who were exposed to three targets performed with a greater detection probability compared with the groups exposed to one and zero targets. The study also concluded that responding may be context and cue-based, as the 0T group had an increase in both hit rate and correct
rejection rate when re-exposed to targets in training environments. Additionally, the one Target group had near equal detection rates when compared with the Three Target group, with no significant differences between them, however, a larger number of false alarms were recorded for Path C compared with the other paths (Porritt et al., 2015).

Edwards et al. (2017) evaluated the influence of target probability with a go/no-go procedure with giant African pouched rats’ by manipulating the proportion of positive and negative TB samples presented to the rat starting with 10% known target probability, which was then systematically decreased in 2% decrements until a 2% target probability was reached. The researchers found that when a 2% target probability was examined, hit rate was much lower compared with the hit rate during the 4% target probability, indicating that decreases in target probability correspond to a decrease in sensitivity. This study further suggests that target probability can be shown to have an effect on responding during a go/no-go procedure, but this experiment was conducted under applied settings where the status of the sample was unknown and therefore not controlled for (Edwards, Ellis, et al., 2017).

Another study which utilised the go/no-go procedure was completed by Haycock (2017) presented hens with the same go/no-go procedure and apparatus as Tashkoff (2017) to assess the influence of target probability on hens. Refer to the Tashkoff (2017) study in Chapter 2, Introduction to understand the apparatus. During experimental sessions, hens were exposed to three conditions, the first condition examined 50%
target probability and 100% reinforcer probability for a hit. The second condition examined 12.5% target probability and the reinforcer probability for hits was decreased from 100% to 75%, then to 50%. During the last condition, the target probability of either 25% or 6% was examined, and similar to the second condition, the reinforcer probability for hits was decreased from 100% to 75%, then to 50%. There was a return to the baseline between each of the three conditions of 100% target and reinforcement probability. Haycock (2017) concluded that hit rate and correct rejection rate were not significantly altered in relation to target probability and stated that the hens showed a bias toward the go response option, as this was the only response option that provided a reinforcement, and so when the reinforcement rate was high, the hens tended to produce a higher rate of hits and false alarms (Haycock, 2017).

Because dogs are so widely used in applied settings, understanding the impact of target probability and its influence on performance is important. Currently, there is little relevant empirical research in regards to go/no-go scent tasks. The current study was conducted under controlled conditions while manipulating the target probability to assess whether target probability alone would influence scent detection accuracy in a go/no-go task. In accordance with the study conducted by Edwards et al. (2017) and Haycock (2017) it was hypothesised that as target probability decreases in a go/no-go scent detection task, hit rate will decrease and correct rejection rate will increase; alternatively, as the target probability increases, hit rate will increase and correct rejection rate will decrease.
Method

Subjects

Four of the same dogs used in Experiment One (Table 2) are used in this experiment with the exception of Ruby as her owners moved and could not bring her in to the laboratory.

Apparatus

The same apparatus used in Experiment One was used in this experiment.

Procedure

Solution preparation

Sample solutions and preparation in this experiment was completed in the same way as in Experiment Two.

Sample preparation

The vials were prepared the same as in Experiment One.

Shaping

Dogs were already trained to use the apparatus, as described in Experiment One (Table 3).

Experimental sessions

The proportion of positive and negative vials used differed from Experiment One. The ratio of positive to negative samples presented to the dogs in a given session was manipulated systematically (Table 6). The criterion to advance to the next ratio was the same as in Experiment One (i.e., six consecutive sessions with hit
rate and correct rejections above 80% with no obvious trend within the last four
data points).

The indication threshold remained constant at 5.5 seconds. Starting with a
baseline of 7:10 (positive:negative) ratio proportion, the proportion of positive
samples were decreased systematically to 5:12, 3:14, and then 1:16. Then a return
to baseline condition (7:10) before positive proportions were increase
systematically to 9:8, 11:6, 13:4 and then 15:2. Unordered conditions were then
examined to 4:13, 12:3, 2:15 and then 6:11. The proportions will now be referred
to as probabilities and are listed in Table 6.

The researchers examined unordered target probabilities to assess
performance at these target probabilities and to test for order effects. The
termination criteria were the same as in Experiment One; 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 session were stopped for
the day. If this occurred for two consecutive days, the dog was withdrawn from
the study. Additionally, if a dog completed a session with hit rate below a value of
0.5 for six consecutive sessions, sessions stopped for that day. If this occurred for
two days in a row, the dog was withdrawn from the study.
Table 6

Target Probabilities Examined

<table>
<thead>
<tr>
<th>Target Probability</th>
<th>Baseline 41% (7:10)</th>
<th>Return to baseline 41% (7:10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29% (5:12)</td>
<td>53% (9:8)</td>
</tr>
<tr>
<td></td>
<td>18% (3:14)</td>
<td>65% (11:6)</td>
</tr>
<tr>
<td></td>
<td>6% (1:16)</td>
<td>76% (13:4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>88% (15:2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23% (4:13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70% (12:3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12% (2:15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35% (6:11)</td>
</tr>
</tbody>
</table>

Data analysis

Data were analysed in the same way as in Experiment One except that target probability was the primary factor instead of indication threshold. The same graphical and statistical analysis completed in Experiment One was completed in this experiment.

Results

Trevor showed high hit rate and correct rejection rate for the duration of the experiment (Figure 19). However, hit rate remained above correct rejection rate
with the exception of when the target probability went back to baseline of 41% where both performance measures were equal, and during 53% target probability where correct rejection rate was above hit rate at a value of 1 and hit rate was at a value of 0.946. During re-examined probabilities, hit rate and correct rejection rate values remained consistent with experimental sessions with a slight upward trend incorrect rejection rate values.

Figure 19. Trevor’s mean hit rate (circles) and correct rejection rate (triangles) for each target probability, phase break indicates when unordered target probabilities were examined.

Tui’s hit rate and correct rejection rate started almost equal with hit rate at 0.913 and correct rejection rate at 0.916, and both performance measures remained high, with correct rejection rate above hit rate values (Figure 20). However, when the target probability was increased to 76% and 88%, correct rejection rate decreased in a downward trend to 0.83 and hit rate increased to a
value of 0.98 from 0.94. During re-examined probabilities, hit rate and correct rejection rate remained consistent with experimental sessions.

![Graph showing hit rate and correct rejection rate for each target probability]

**Figure 20.** Tui’s mean hit rate (circles) and correct rejection rate (triangles) for each target probability, phase break indicate when unordered target probabilities were examined.

Katie displayed high levels of both hit rate and correct rejection rate throughout the experiment, with the exception of correct rejection rate at the target probability of 6% which decreased to a value of 0.88 (Figure 21). Correct rejection rate surpassed hit rate at the target probability of 53% and 65%. When targets were re-examined, values for both performance measures remained high between the values of 0.92 and 1.
**Figure 21.** Katie’s mean hit rate (circles) and correct rejection rate (triangles) for each target probability, phase break indicate when unordered target probabilities were examined.

Ella started off with hit rate lower than correct rejection rate during baseline (Figure 22). Hit rate increased and remained above correct rejection rate until the target probabilities of 65% and 76% were examined. Hit rate then decreased from 0.98 to 0.69, but then recovered to 0.90 to equal correct rejection rate during 88% target probability. During re-examined target probabilities, there was a slight variation in performance, but overall performance remained high.
Figure 22. Ella’s mean hit rate (circles) and correct rejection rate (triangles) for each target probability, phase break indicate when unordered target probabilities were examined.

Hit rate and correct rejection rate for all dogs were combined to assess differences across dogs (Figure 23). For the majority of the experimental sessions, both performance measures remained at a high level starting with a value of 0.95 for both measures. Correct rejection rate decreased slightly on the 6% target probability but rose back up to a value of 0.98 to surpass hit rate values when the target probability returned to baseline of 41% and at 53% and 65%. Correct rejection rate then decreased slightly in equal value to hit rate, then below the hit rate value for the 88% target probability. During reviewed target probabilities performance values varied slightly but remained high between 0.9 and 0.99 value.
Figure 23. All dogs mean hit rate (circles) and correct rejection rate (triangles) for each target probability, phase break indicate when unordered target probabilities were examined.

The mean hit rate levels for all dogs for each target probability is represented by the bars in Figure 24. All dogs tended to perform above the mean with Ella performing under the mean during the 56%, 76%, and 70%. For all dogs, hit rate remained between 0.9 and 1. However, there was a slight drop below 0.9 when the target probability was examined at 65%. During the re-examined probabilities, all dogs performed above the mean, except for Ella on 70% and Tui on 12%.

Correct rejection rate values appear to show idiosyncratic differences between dogs but the overall mean correct rejection rate values for all dogs at each target probability remained high throughout the experimental session (Figure 25). Katie tended to be equal to or above the mean, where the other dogs tended to be equal to or below the mean. Overall, there was a slight decrease in mean
correct rejection rate when 6% probability was examined. A slight downward trend from 65% to 88% target probability was observed for the mean correct rejection rate for all dogs. During the re-examined conditions, there is an upward trend for the mean correct rejection rate for all dogs.
Figure 24. Mean hit rate summary data. The mean hit rate across all dogs for each indication threshold is also included (black line).
Figure 25. Mean correct rejection summary data. The mean hit rate across all dogs for each indication threshold is also included (black line).
ROC curves for individual dogs are presented indicating that all dogs performed with high accuracy (Figure 26). The ROC curve for all dogs collectively indicates that mean accuracy was high in all conditions (Figure 27).

*Figure 26. ROC curves for individual dogs.*
A Shapiro-Wilk test of normality was conducted which revealed that the hit rates in the first four target probabilities (41%, 29%, 18% and 6%; Table 7) were not normally distributed. A non-parametric Friedman test of differences among repeated measures was conducted and rendered a Chi-square value of $X^2(8) = 11.523$, $p = .174$ which was not significant ($p>.05$).

A visual inspection of the linear regression shows a slight downward trend between increased target probability and hit rate (Figure 28). A Durbin Watson statistic of 1.893 determined independence among the residuals. A visual inspection of the standardized residuals scatters plot indicates that there may be some variance in the residuals and so a weighted squares test was completed confirming homoscedasticity.
Table 7

Shapiro-Wilk Test of Normality for Hit Rate Using Summary Data

<table>
<thead>
<tr>
<th>Target probability</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>41%</td>
<td>.001*</td>
</tr>
<tr>
<td>29%</td>
<td>.000*</td>
</tr>
<tr>
<td>18%</td>
<td>.001*</td>
</tr>
<tr>
<td>6%</td>
<td>.000*</td>
</tr>
<tr>
<td>41%</td>
<td>.272</td>
</tr>
<tr>
<td>53%</td>
<td>.056</td>
</tr>
<tr>
<td>65%</td>
<td>.062</td>
</tr>
<tr>
<td>76%</td>
<td>.439</td>
</tr>
<tr>
<td>88%</td>
<td>.272</td>
</tr>
</tbody>
</table>

Note. Findings were considered significant at the $p<.05$ level. Significant findings are marked with an asterisk. Indication threshold was not found to significantly predict correct rejection rate and a negative slope was found ($\beta = -.173$, $t = -1.025$, $p = .312$, 95% CI[-.001, .001]). This yielded a predictive model for specificity performance ($S = 0.9854 + -0.0329* \text{Target probability}$).
Figure 28. Linear regression analysis of hit rate for all target probabilities examined for dogs individually (empty black) and the mean (full black). Probabilities assessed for order effects (triangles) were also included.

A Shapiro Wilk test for correct rejection rate was then completed to test normality for target probability which resulted in three target probabilities, not within normal limits (41% revisited, 53%, 65%; Table 8). A non-parametric Friedman’s test of differences among repeated measures was conducted and rendered a Chi-square value of $X^2(8) = 7.707, p = .463$ which was not statistically significant.

On visual analysis, of the linear regression (Figure 28) there appears to be a slight upward relation between increase target probability and correct rejection rate. A Durbin-Watson statistic of 0.373 was calculated which indicated that the residuals are not independent. A simple linear regression was calculated to predict correct rejection rate based on target probability.
Table 8

*Shapiro-Wilk Test of Normality for Correct Rejection Rate Using Summary Data*

<table>
<thead>
<tr>
<th>Target probability</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>41%</td>
<td>.161</td>
</tr>
<tr>
<td>29%</td>
<td>.161</td>
</tr>
<tr>
<td>18%</td>
<td>.729</td>
</tr>
<tr>
<td>6%</td>
<td>.705</td>
</tr>
<tr>
<td>41%</td>
<td>.001*</td>
</tr>
<tr>
<td>53%</td>
<td>.001*</td>
</tr>
<tr>
<td>65%</td>
<td>.001*</td>
</tr>
<tr>
<td>76%</td>
<td>.272</td>
</tr>
<tr>
<td>88%</td>
<td>.272</td>
</tr>
</tbody>
</table>

*Note.* Findings were considered significant at the $p<.05$ level. Significant findings are marked with an asterisk.

A significant regression equation was found ($F(1,50)= .036$, $p= .849$), with an $R^2$ of .000. The standardised scatterplot of residuals was examined which indicated a linear trend which suggested to violate the assumption of homoscedasticity. A weighted least squares analysis confirmed homoscedasticity. Indication threshold was not found to significantly predict correct rejection rate and a positive slope was found ($\beta= .027$, $t=.191$, $p = .849$, 95% CI[-.001, .001]). This yielded a predictive model for specificity performance ($S = 0.9543 + 0.0057* Target probability)$)
Figure 29. Linear regression analysis of correct rejection rate for all target probabilities examined for dogs individually (empty black) and the mean (full black). Probabilities assessed for order effects (triangles) were also included.

**Discussion**

The aim of this experiment was to investigate the influence target probability had on hit rate and correct rejection rate during a go/no-go scent detection task when the proportion of positive and negative samples were manipulated systematically. Overall, there was no significant finding when a statistical analysis was completed, however a medium effect was found between target probability and hit rate, and a small effect between target probability and correct rejection rate. Specificity dropped for most dogs, with the exception of Tui, during the lower target probabilities (6%) and then increased above sensitivity during the higher target probabilities (Figure 22). The results from the current experiment were not statistically significant, and based on a visual analysis on the graphical analysis, hit and correct rejection rate are not altered by target probability manipulations.
For example, Trevor’s hit rate was high through all target probabilities examined, with hit rate below correct rejection rate. However, hit rate did drop below correct rejection rate during the 6% target probability examined. This indicates that the influence of target probability had a small effect on Trevor’s performance which was confirmed in the statistical analysis. Similarly, target probability had a small effect on Tui’s performance, although he had a larger correct rejection rate than hit rate. This indicates that as target probability increased, correct rejections increased and hit rate decreased. During the examination of the smaller target probabilities,

The results from this study are similar to the study by McCarthy and Davison (1979), who also found that target probability had little influence on responding. That study involved experiments in which target probability alone was manipulated and experiments where both target probability and reinforcement probability were manipulated in a yes-no procedure. The researchers concluded that reinforcer probability had more of an influence on hit rate and correct rejections, while target probability had little influence on performance. However, the current study adopted the go/no-go procedure where a decrease in target probability corresponds to a decrease in reinforcer probability. Therefore, the comparison of the current study and the study from McCarthy and Davison (1979) indicate that when an alternative reinforced response is available different results may be obtained and it is not clear what the outcome with a go/no-go procedure will be because of this difference in procedure. However, from the results from the current study, the outcome shows the same finding, as McCarthy and Davison as only small effects were found (McCarthy & Davison, 1979). However, Haycock (2017), who used the go/no-go procedure with hens on a visual
discrimination task, found similar results to McCarthy and Davison (1979) as the target probability also had a little effect on the chickens in her study using a go/no-go procedure suggesting that target probability has little impact on scent detection performance in a go/no-go scent detection task within the target probability values examined (Haycock, 2017).

Also in agreement with the present study, Gazit et al (2005) found an increase in sensitivity when low target probability conditions were examined, compared with higher target probability conditions, however, Gazit and colleagues’ (2005) study differed to the current study in many respects. In addition, the results from Porritt and colleagues’ (2015) study indicated that a group of dogs exposed to one target had equal detection rates to a group exposed to three targets with no significant differences between the two. However, a larger number of false alarm responses were recorded from the one target group. A larger number of false alarms and a decrease in correct rejections suggest that the one target group had a decrease in correct rejections when re-exposed to search areas with four targets, which is in line with the results from the present study. However, that study also had a procedure that differed in many respects which may show different results in regards to the go/no-go procedure (Porritt et al., 2015). In addition, Edwards and his colleagues (2017) also obtained similar results, as target probability had little effect on accuracy. However, that study examined lower target probabilities that the current study did not examine which showed different results. Future research examining the effects of lower target probabilities using the same procedure as in the
current study might yield more significant results than observed in the current study (Edwards, Ellis, et al., 2017).

A limitation of this study includes the number of target probabilities examined may have been too small, and examining a larger range of target probabilities may have been more useful to further examine the influence on performance. Future research in this area should increase the range of conditions examined. Additionally, when the probability of the positive samples decreased below 18% accuracy declined rapidly, to reduce the potential influence of pervious residue or volatiles, during the examination of target probabilities below 18%, all segments containing positive samples previously were wiped with alcohol, they start to indicate where a positive sample previously was. This was only done for low target probabilities and was not completed for all target probabilities examined. Different results may have been found if all segments had been wiped.

The results from the current study indicate that as the target probability increases, the number of hits (sensitivity) decreased and the number of correct rejections (specificity) increased. This indicates that the matching law failed to predict responding when targeting probability in a go/no-go scent detection task with dogs increased. However, the findings from this study should make an important contribution to research on the go/no-go procedure and target probability, similar results are obtained with dogs in a scent detection task as those obtained with other species, sense modalities and procedures. The results from the current study suggest that high accuracy can be obtained across a broad range of target probabilities, which suggests that the future applied research or
operations with the apparatus will not be adversely influenced by high or low proportion of target samples and, thus it will be useful for future canine scent-detection research.
Chapter 4: General Discussion

Together, the results from the two experiments undertaken in this study provide important information for further scent detection studies with dogs. The first experiment highlighted the influence of response requirement on responding, as an increase in the length of the indication threshold increased the correct rejection rate and decreased the hit rate as the longer indication threshold became more effortful than the activation of the limit switch. The results from Experiment One tended to be in agreement with other studies investigating response effort, which have also found that an increase in response thresholds decrease hit rate and increase correct rejection rate. The results from Experiment One also suggest that a dog’s indication threshold should be individually assessed as there was variation in where over accuracy was the highest, and where accuracy started to deteriorate. Nevertheless, there did appear to be indication thresholds which tend to produce high performance for all dogs which could serve as default values until individual assessment is conducted, or if individual assessment is not practical.

During Experiment Two, target probability had a small influence on accuracy with a small effect as no significant effects were observed. But a small increase in hit rate was observed when target probability was low with a small decrease in correct rejections and alternatively, as the target probability increased, hit rate decreased and correct rejection rate increased. The results from the current study were in line with the studies investigating the effects of target probability in go/no-go signal detection tasks with other animals.

To increase internal validity of the two experiments outlined in this study, each dog performed the experiments in the room alone to reduce any human
cuing or bias. This also removed any unintentional reinforcement from human behaviour. Dogs cannot work alone unless something is programming sample presentation, delivery of reinforcement and providing auditory feedback. In the current study, an automated apparatus controlled with a developed software system, presented the samples to the dog and delivered reinforcement. Both the automation of the procedure and removal of human cueing reduced any human error to ensure that the experiments were a true test of the hypothesis. These aspects have also been discussed by Edwards and colleagues (2017) who developed guidelines on best practice when employing dogs to detect human diseases in a scent detection task (Edwards, Browne, et al., 2017). Additional points from those guidelines describe the importance of replicating training settings to match experiences an animal will experience in operational settings. The go/no-go achieves this by only reinforcing correct indications in the presence of the target, and not reinforcing false alarms, misses or correct rejections, as these are not reinforced as much as hits (Edwards, Browne, et al., 2017).

Another set of scent detection research guidelines were developed by Johnen, Heuwieser and Fischer-Tenhagen (2017). The researchers reviewed literature on scent detection with dogs, and compiled important factors which should be considered for best practice. The guidelines indicated best practice methods for study design, dog training and handling procedures and all studies reviewed were rated according to a 10-point scale. Studies had to achieve a score of 6 or above to be considered a successful study according to the authors’ criteria. Factors considered
when rating studies included sample size of the detector dogs, training methods, the number of different samples used during training and testing, the randomization of samples, blinding of handlers, presentation of results, and critical and objective discussion of results. The researchers examined 54 studies and found that 40 of those studies had a score above 6 on the 10-point scale (Johnen, Heuwieser, & Fischer-Tenhagen, 2017).

The current study meets criteria of the factors outlined by Johnen et al. (2017). For example, the current study used amyl acetate as the target scent, which is a scent commonly used in the field of scent detection. The current study also required a decision response of either yes or no, and the current study used the go/no-go procedure which required the dog to perform a go or no-go response which is in line with this factor from the authors. The number of samples used, the use of negative controls, and the randomization of samples were all suggestions from the guidelines, which in line with the procedure of the current study. Another recommendation of the guidelines was to use a scent that was novel to the dog; as far as the researcher knew, no dog had prior contact with amyl acetate (although the odour resembles the smell of bananas, which the dogs may have had contact with before). In addition, all measures were taken to ensure no cross contamination occurred between positive and negative samples. All dogs performed the scent task alone in the absence of the experimenter which is a suggestion from the authors. Therefore, double blinded handlers were not required for either experiment in this current research. Although, an integrity check of all procedures were completed to ensure all procedures were completed as stated to avoid any cross-contamination. Finally, the authors suggest that the training methods are detailed; the current study has provided a detailed explanation of the
shaping procedure (Table 3) and SOP’s (Appendix B), but the dogs were trained by an inexperienced trainer but guidance was provided from experts (Johnen et al., 2017).

There were some limitations to this study. First variables that may have influenced responding included; apparatus breakdowns and updates, the feeder malfunctioned, people walking past the laboratory and noise outside of the laboratory, illness of dogs, and other dogs barking outside the experimental room may have encourage the dog to engage in off task behaviour which may have influenced his responses during the scent detection task. Additionally, the type of kibble used as the reinforcer was changed occasionally due to supply which may have subsequently influenced responding due to a higher preference to the novel food. However this is only an assumption. But anecdotally it was observed that all dogs consumed all kibble types readily, and thus it is thought that the impact of this is minor.

Possible future research could examine the effects of using different targets which are commonly used in applied settings and see if different results would be observed in hit rate and accuracy. It is quite common to use dogs for scent detection in various applications. The proposed research has the potential to help us understand how we should arrange training and operations in order to maximize the performance of scent detection dogs under these operational conditions. This study advance our knowledge of the go/no-go procedure with scent-detection dogs. Trainers and handlers should consider the influence of indication
threshold when working with dogs as it may impact accuracy in applied and laboratory settings. Furthermore, the results from this study will benefit future studies completed at the dog lab at the University of Waikato utilising the scent-detection apparatus for basic and applied scent-detection research.
References


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


Appendix A: Dog Initial Screening Form

Thank you for your interest in our dog behaviour research. We are looking for dogs who enjoy going to new places and meeting new people – and who really like working for food. We have some other criteria for potential research participants, so if you are interested in your dog possibly taking part, please provide the following information.

Is your dog fully vaccinated (standard vaccines: distemper, hepatitis, parvovirus)?  
Yes / No
If no, please explain briefly:
________________________________________________________________________
________________________________________________________________________

Does your dog enjoy meeting new people?  
Yes / No
E.g., are they friendly and comfortable around strangers?
If no, please explain briefly:
________________________________________________________________________

Is your dog comfortable being handled by other people?  
Yes / No
E.g., is your dog happy to be touched on their body, neck, head, tail, paws, etc.? If no, please explain briefly:
________________________________________________________________________

Is your dog comfortable going to new places?  
Yes / No
E.g., is your dog relaxed and happy (showing no signs of stress) when you go somewhere new? If no, please explain briefly:
________________________________________________________________________

Is your dog comfortable when you leave them, including at home alone and new places?  
Yes / No
E.g., is your dog relaxed and happy (showing no signs of stress) when you leave them? (Dogs will not be left alone at our training facility, but we would like to know if they might have any separation-type anxieties.)
If no, please explain briefly:
________________________________________________________________________
Does your dog like working for food?  
Yes / No
If no, please explain briefly:
______________________________________________________________________________________________

Can your dog eat any food, including kibble (biscuits) and different kinds of meat products?  
Yes / No
If no, please explain briefly:
______________________________________________________________________________________________

Is your dog comfortable with people getting near their food?  
Yes / No
E.g., if your dog has shown any aggression (freezing, growling, snarling, biting) around food, please select ‘no’.
If no, please explain briefly:
______________________________________________________________________________________________

Is your dog friendly towards other dogs?  
Yes / No
E.g., if your dog has shown any aggression or fear towards other dogs, please select ‘no’. (We will not necessarily have more than one dog at the training facility at once. If we do, it will be with permission of all owners and the dogs will be contained separately.)
If no, please explain briefly:
______________________________________________________________________________________________

Is your dog comfortable with unexpected/loud noises, such as beeping sounds?  
Yes / No
If no, please explain briefly:
______________________________________________________________________________________________

Is your dog free of medical conditions that could be aggrevated by repetitive walking?  
Yes / No
E.g., if your dog has any joint or other problems that might be affected, please select ‘no’.
If no, please explain briefly:
______________________________________________________________________________________________

Would you be able to drop off and pick up your dog in the morning/afternoon so that your dogs spent just half a day with us (our facility is at the University of Waikato main campus)?  
Yes / No
Please indicate which times are more convenient:
__________________________________________________________________
_________________________________

We want to make sure that all dogs enjoy participating in our research. If you answered “no” to any of these questions, this may indicate that your dog is not suitable for some of this research; however, it does not necessarily exclude them from taking part. A researcher will be in touch with you to discuss the information you have provided here. Thank you for taking the time to complete this form.

Please email this form to: claudia.giezen@outlook.com
Appendix B: Experiment One SOP

Start of the session.

- Load up the vials according to the novel sequence generated.
- Open up the detection application on the computer and type in the number of corresponding to the dog currently used in the subject box.
- Check the novel sequence in the application appropriately matches the randomly generated sequence developed.
- Ensure the indication threshold to be examined is correct.
- Turn on all video recording devices by pressing the red button on the left, right and sound recording screens.
- Take the dog out of the kennel and into the experiment room.
- Push the flap once to ensure it is working.
- Close the door and try not to make any noise while the dog is completing the experiment.
- Start the experiment.
- Record data by hand.

Changing sequence.

- To ensure that the dog is exposed to a novel sequence each time, the sequence will need to be changed often.
- With fresh gloves, take off the lid from the apparatus.
- Place the red bucket in the middle of the apparatus. Take out all 17 samples individually and placing all positive samples in the red bucket (to
prevent any cross contamination in case of spillage) and the negative samples outside of the bucket.

- If there is a spillage of either negative or positive samples, take away the spilled vial and any vial that may have come into contact with the sample. Take the sticker off these vials and throw it away. Empty any remaining sample into the rubbish chemical jar. Place the actual vial into a separate bucket and create a new sample (refer to “refresh samples” in this SOP for help).

- Referring to the next sequence pre generated, first place all the negatives into their appropriate segments, then the positives.

- Place apparatus lid back on.

- Wait at least 5 minutes before starting the next experiment, to let the sample volatiles fill the segment.

Increasing the indication threshold

- Once the dog has completed six sessions, look at the raw data to see if there is a positive or negate trend.

- If there is a positive or negative trend, complete more sessions at the current indication threshold until the data is stable.

- Once stable, increase the indication threshold by 0.5 seconds. Complete the above steps of the SOP for the next indication threshold.

Refreshing Samples.

- Samples need to be refreshed after 2 hours have elapsed since the current samples have been created.
• With new gloves take out all vials, place all positives in the red bucket and all negative samples outside the bucket in the middle of the apparatus.

• Referring to the next sequence, place all negative samples into their appropriate segments.

• Take the red bucket, which still has all the positive samples, to the prepping station.

• Empty the sample from each vial into the “rubbish chemical jar”

• Using the same vials, place 2 mL of the positive sample solution using the autopete with fresh gloves.

Off task behaviour.

• Off task behaviour in measured in the amount of time the dog spends away from the observation response, including the lever.

• If the dog has taken longer than two minutes from taking his nose out of the sample port to placing his nose back in again, we class this as off task behaviour.

• After two minutes have elapsed, give a level one prompt but still remain outside of the experimental room to make the apparatus beep on the computer application at least 20 times or until they start the observation response.

• If the level one prompt did not work, then use the level two prompt of slightly opening the door of the experimental room.

• If neither prompted has been success in getting the dog to re-engage with the apparatus, then terminate the session.
• Terminate the session means to take the dog out of the room, don’t record any data collected for that session.

• Give the dog a walk, as potential reason for off task behaviour may include a full bladder or bowl. Then start the session again.

• If the dog engages in any off task behaviour throughout 50% of the sessions for a given indication threshold, two days in a row. Then terminate them from the study completely.