Journal of Veterinary Behavior 13 (2016) 34–38

Contents lists available at ScienceDirect

Journal of Veterinary Behavior

journal homepage: www.journalvetbehavior.com

Short Communication

Wildlife detection dog training: A case study on achieving generalization between target odor variations while retaining specificity

Cor Oldenburg Jr.[a,b,*], Adee Schoon[c], Ignas M.A. Heitkönig[a]

[a] Resource Ecology Group, Wageningen University, Wageningen, The Netherlands
[b] Speurhonden Instructie School, Beverwijk, The Netherlands
[c] Animal Detection Consultancy, Vorchten, The Netherlands

ARTICLE INFO

Article history:
Received 25 January 2016
Received in revised form
9 March 2016
Accepted 25 March 2016
Available online 6 April 2016

Keywords:
canine
odor
detection
scat
Eurasian otter
conservation

ABSTRACT

Wildlife detection dogs are required to correctly discriminate target wildlife species odor from nontarget species odors (specificity), while enabling some degree of target odor variation (generality). Because there is no standardized training protocol, and little knowledge on training efficiency, we conducted a case study to test a dog's training efficiency in detecting 2-week-old wild otter (*Lutra lutra*) feces (spraint) odor among feces odors from 6 other large mammal species that often share the otter's natural habitat, including fox (*Vulpes vulpes*), hare (*Lepus europaeus*), roe deer (*Capreolus capreolus*), and cattle (*Bos taurus*). The dog was trained using a standard multiple-choice carrousel in a stepwise protocol. We started with odor samples from fresh captive otter spraints and progressed toward 2-week-old spraints from wild otters among other mammalian dung odors and tested for specificity and generality after each training step. We show that training on only 2 variations of spraints from captive otters enabled the dog to detect all desired spraint odor variations in our protocol, indicating a rapid generalization to variations of spraint odor the dog was not trained on, while retaining specificity. Testing such concept formation of target odors should be included in detection dog training and certification and could serve as a quality control measure of detection dog performance.

Introduction

Wildlife detection dogs can be trained to identify traces of target animals and their scats (de Oliveira et al., 2012; Wasser et al., 2012; Browne et al., 2015). The dogs are often required to correctly discriminate target wildlife species dung odor from nontarget species dung odors within the same habitat (specificity), while enabling some degree of target odor variation such as caused by diet or age of the target species (generality). The goal of the training of wildlife detection dogs is concept formation: the dogs learn to respond to odors similar to those used in training, based on recognizing common odor components of the particular target wildlife species. Although this is often called generalization (Oxley and Waggoner, 2009; Lazarowski and Dorman, 2014), strictly speaking generalization is an accidental outcome of a learning situation [the tendency for stimuli similar to the original stimulus to produce the response originally acquired (Ghirlanda and Enquist, 2003; Reber et al., 2009)], and not the intended outcome. Concept formation, leading to identical responses to different sources or variations of target odor, is often assumed (Cablk and Heaton, 2006), but our literature review did not reveal a systematic assessment to check detection dogs on this point during their training. The few published studies in this field indicate (1) that dogs do not respond to all desired target odor variations if there is insufficient variety in training aids (Oxley and Waggoner, 2009) and (2) that several variations of target odor training are necessary to enhance generalization and thus promote the formation of a concept (Fischer-Tenhagen et al., 2011; Oxley and Waggoner, 2009). However, a step-by-step assessment of this process has not been conducted before.

Our study was designed to systematically train a detection dog to discriminate an increasingly complex set of target odor variations from decoy odors from scats of 6 other mammal species and to

* Address for reprint requests and correspondence: Resource Ecology Group, Wageningen University, Vrijzicht 27, 1068 CD Amsterdam, The Netherlands. Tel: +316 305 929 58.
E-mail address: oldenburgcj@gmail.com (C. Oldenburg).

http://dx.doi.org/10.1016/j.jveb.2016.03.008
1558-7878/

systematically assess its tendency to generalize target odors. The goal of the training was for the dog to respond to a spraint (otter scat) as it would be found in nature: from wild otters, slightly degraded through outdoor aging. We first trained the dog on fresh captive otter spraints and tested its response to 3 scat age levels (from fresh, through aged 1 week to aged 2 weeks) and then to diet at 2 levels (from captive to wild) combined with 3 age levels, that is, 6 target scent variations in total. After each step of the training, specificity was tested against decoy samples, and the level of generalization was tested by offering unfamiliar spraint samples.

Methodology

Design, dog, target odors, training carrousel, and responses

The study was designed to train the dog progressively to sniff out 2-week-old target odors amidst decoy odors and to test progress at each step (Figure 1). The dog's tendency to generalize was assessed in 3 tests. During training with each target odor variation, test 1 was conducted to see if the dog generalized to a similar spraint of the same odor variation but from another animal. If yes, test 2 was conducted to see if she could detect 2-week-old spraint from a wild otter. If so, test 3 was carried out to confirm concept formation by including all available target and decoy odors. On failing a test, the training was picked up at the point where it had failed.

We used a female Malinois of 12 years old, keen to hunt and search for detecting objects with human scent in the field. Spraints (scats) from 3 groups of seafood-fed captive otters (housed separately) were collected in Natuurpark Lelystad, a Dutch nature education facility. Wild otter spraints were collected on multiple locations in the Dutch nature reserve Weerribben-Wieden, where animals consumed fresh water fish (Leonards et al., 1997). Like in a scent-discrimination experiment, we used a 6-arm multiple-choice carrousel (Schoon and Haak, 2002; Schoon et al., 2014) in a training facility to train and test the dog, making use of blanks, target, and decoy samples. The dog's response to each and every odor sample or blank in the carrousel was individually recorded and classified in a 2 × 2 confusion matrix as either true positive (alert, target odor present), true negative (no alert, target odor not present), false positive (alert, target odor not present), or false negative (no alert, target odor present).

Decoy odors and preparation of samples

Six feces odors were chosen as decoys during training and testing, selected for their occurrence in Dutch nature reserves and their tendency to distract dogs. They were feces of fox (*Vulpes vulpes*), rabbit (*Oryctolagus cuniculus*), hare (*Lepus europaeus*), roe deer (*Capreolus capreolus*), cattle (*Bos taurus*), and horse (*Equus ferus caballus*). For each species, feces samples were collected from multiple locations to ensure variability. Like the target odor, the decoy odors varied in age; fresh, 1-week-old and 2-week-old samples were available for each decoy odor.

All fecal samples for target and decoy odors were collected and handled with plastic gloves. Only fresh samples were collected, stored in plastic containers, individually labeled, and kept in clean trays in a freezer, separated by species. A number of fecal samples of each species were stored separately outdoors underneath a shelter for aging 1 week or 2 weeks, respectively, on which they were frozen and stored too. Frozen feces samples were cut into small, thin slices of approximately 1 cm^2 before training or testing. This enabled reuse of the same frozen sample without additional degradation and contamination issues (*cf.* Browne et al., 2015 and

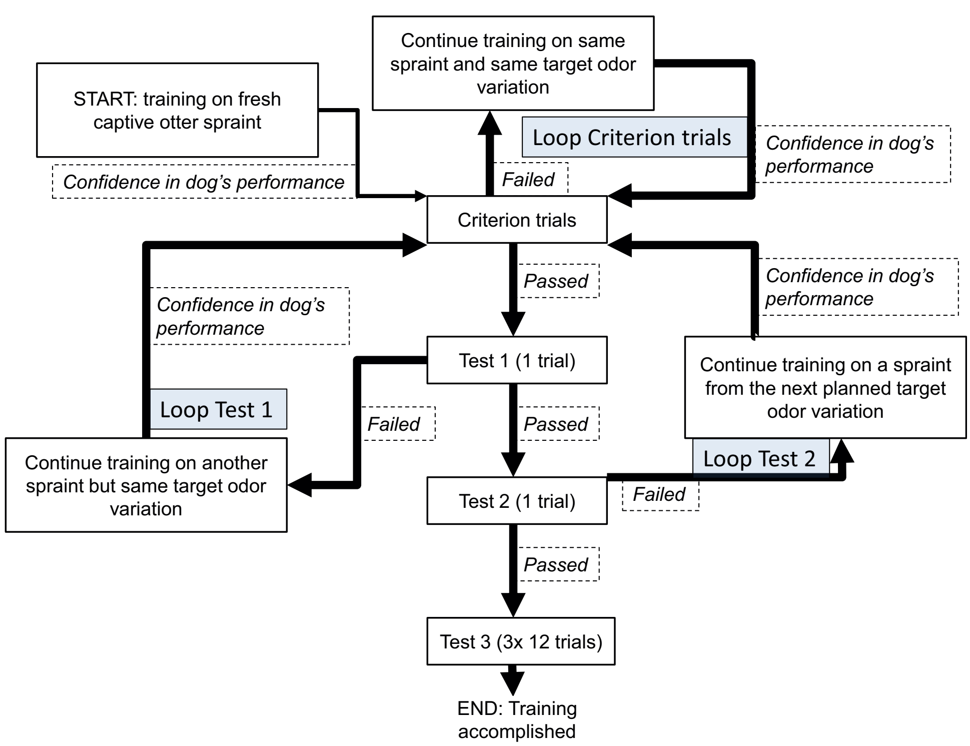

Figure 1. Flowchart of the designed schedule for training and testing detection dog generality and specificity. Training started with fresh captive otter spraint. Criterion trials were done if there was confidence in the dog's performance. If the dog did not pass the 5/6 criterion, training continued with the same spraint of the same target odor variation (loop criterion trails). If the 5/6 criterion was passed, the experiment continued with test 1. If test 1 was not successful, training continued with a different spraint from the same target odor variation (loop test 1). If test 1 was successful, the experiment continued with test 2. If test 2 was not successful, training continued with a spraint from the next planned target odor variation (loop test 2). If test 2 was successful, the experiment concluded with test 3. The sequence of target odor variations was as follows: fresh captive otter spraint > 1-week-old captive otter spraint > 2-week-old captive otter spraint > fresh wild otter spraint > 1-week-old wild otter spraint > 2-week-old wild otter spraint.

references therein). Care was taken to prevent odor contamination by using clean equipment at all times.

Pretraining routine

The dog was pretrained to check the 6 sample locations on the carrousel with either blank samples, a target odor, or—later—the decoys and to perform a sit-alert when it encountered a target (spraint) sample. Each event in which the dog checked the randomly placed odor samples on the carrousel was called a “trial” (Fjellanger et al., 2002), which started with a search command and ended either with the dog performing an alert (correctly or incorrectly), or not alerting to any checked sample. The dog was called by the handler and rewarded with food only if she correctly alerted for a target odor sample. Up to 20 trials per day were carried out, with no more than 5 trials in a session, separated by breaks to prevent fatigue. After each trial, all samples were removed. During pretraining, only the target odor was present at first (captive fresh), along with blank samples. The decoy odors were introduced later on, adding them randomly when the handler was confident about the dog’s performance. Pretraining was considered completed when the dog worked proficiently with the carrousel, and all decoy odors had been presented and correctly rejected in the setup.

Training protocol and the 5/6 criterion trials

After pretraining, the training and testing protocol was carried out according to the flowchart (Figure 1). During training, only one sample of a target odor variation was used in target trials, along with 5 decoy odors of the same age, but no blanks. Decoy odor samples collected from a different location were used when another spraint sample was used. In “zero trials,” only decoy samples were available and the handler determined the setup. Once the handler was confident about the dog’s performance, a set of 6 “criterion trials” was conducted to confirm the dog’s performance at that level, where another person placed the odors in the carrousel, to which the handler was blind. Criterion trials continued until the dog alerted to the target odor in at least 5 of 6 trials (the 5/6 criterion)—followed by test 1—or until the handler decided to return to training in case the dog failed (Figure 1: loop criterion trials).

Testing routines and protocol

Test 1 was conducted after the dog passed the criterion trials (Figure 1) to check if the dog generalized to a spraint from another individual while retaining specificity. It consisted of a single trial with decoy samples from a different location and a similar spraint as used in training at that point (but from a different otter group). If the dog did not alert to the target odor, the test was classified as not successful, after which training continued with a new target sample, from a second individual (Figure 1: loop test 1). If test 1 was successful, test 2 was conducted, also consisting of a single trial, with a 2-week-old spraint from a wild otter, along with 2-week-old decoy samples. Test 2 thus represented a simple version of the final objective of the training and was done each time when test 1 was successful. If the dog did not alert to the target sample in test 2, the test failed, and training resumed with the next step (Figure 1: loop test 2). If test 2 was passed, test 3 assessed the dog’s concept formation by checking her response to all available variations of the target. Test 3 consisted of 3 series of 12 trials, of which 6 target trials and 6 zero trials. Every target odor (captive or fresh, each with 3 age variations) was included only once in the setup for the target trials. All tests were conducted in a blind fashion, that is, presence and position of the targets were unknown to the handler. During tests 1 and 2, one arm carried the target odor sample, whereas the other arms carried one of the six decoy odors (with a maximum of one sample per decoy species). During test 3, every arm of the carrousel carried 1 of the 6 decoy odors (with a maximum of one sample per decoy species per trial), but in target trials, one of the decoy odors was replaced by spraint odor. Dog handling and evaluation during tests were the same as in training.

Data analysis

All responses during the trials were entered in a confusion matrix (Akay, 2009). For further analyses, only the responses during the blind trials (handler unaware of sample locations on carrousel) were used. Separate confusion matrices were made for criterion trials and test 3. To determine whether there was an association between the presence of target odor and the performance of an alert, the values for the number of true positives, true negatives, false positives, and false negatives in the confusion matrices were analyzed with a Fisher exact test.

Results

After 412 training trials (of which 302 trials in pretraining) with fresh captive otter spraint and meeting the 5/6 criterion, test 1 revealed that the dog immediately responded to a fresh spraint from a captive individual from a different group (Figure 2). In test 2, however, she did not detect the goal spraint (wild, aged 2 weeks) based on this training, so the training continued by including a spraint from the next planned target odor variation, that is, adding a 1-week-old spraint from a captive otter to the group of target odors

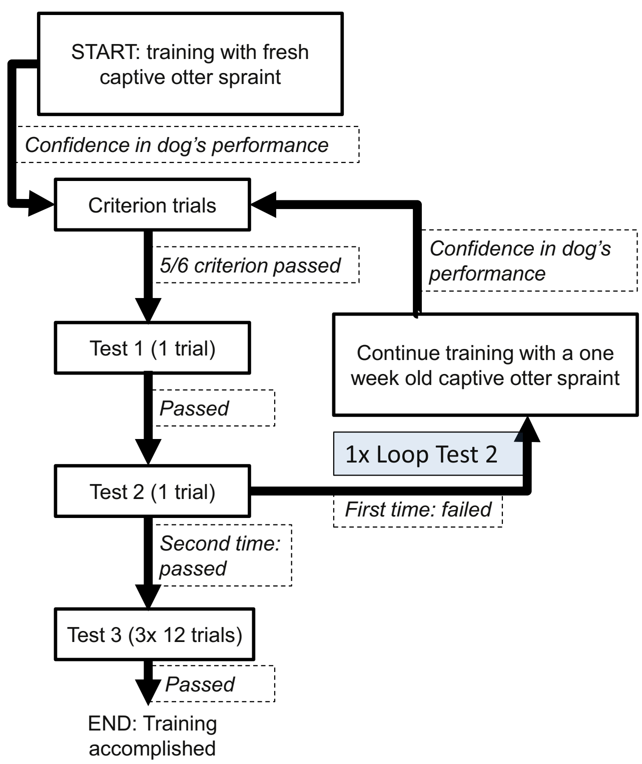

Figure 2. Flowchart of the results of the training and testing of the detection dog in this study. After training on fresh captive otter spraint, the dog passed the criterion trials and subsequently test 1 too but she failed test 2 for the first time, producing a false negative. The training thus continued with loop test 2 on a 1-week-old captive otter spraint. Again, criterion trials were performed when there was confidence in the dog’s performance. After passing the 5/6 criterion, she also passed test 1 and test 2. The study was concluded with test 3, which consisted of 3 series of 12 trials.

in the loop of Test 2 (Figure 2). After an additional 117 training trials and meeting the 5/6 criterion for the 1-week-old spraint, the dog was tested again with Test 1, this time using a 1-week-old spraint from a new captive individual. Again, she immediately responded and subsequently underwent test 2, now for the second time, in which she scored a true positive for the target spraint (wild, aged 2 weeks). There was a significant association between an alert and the presence of target odor during the criterion trials (Fisher exact test, $P < 0.0001$). The confusion matrix for these trials is presented in Table 1.

Test 3 was performed to further assess whether the dog indeed had formed a generalized concept of all target odor variations and discriminated against the decoys. A significant association between an alert and the presence of target odor was found in Test 3 (Fisher exact test, $P < 0.01$). The confusion matrix for test 3 is presented in Table 2. Table 3 summarizes the dog's responses to the target odor variations during the first and second repeat series of 12 trials each. In the first series of 12 trials, the dog performed a true positive for all target odor variations, except for the 1-week-old wild otter spraint, due to a false positive for a sample earlier in the setup. Also during the second and third series of 12 trials, the dog performed a true positive for all target odor variations, except when she gave a false positive alert for a sample earlier in the carrousel. These results confirmed that the dog was able to detect spraint odor in general, because the training on the 2 variations of spraints led to a systematic response to all variations in test 3: every variation was detected at least once, as can be seen in Table 3. Moreover, 16 of the 18 zero runs conducted as part of this test were performed without eliciting any response of the dog, confirming that the dog discriminated against the decoy odors.

Discussion

Although this study involved just 1 dog and 1 target odor, it illustrates a number of key aspects of generalization and concept formation in detection dog training. This study confirms the assumption that trained detection dogs do generalize between different sources or variations of target odor. The dog responded to the 6 desired variations of target odor after being trained on only a 2 of them, in contrast to the study by Oxley and Waggoner (2009) who found that training on 4 variations of target odor (gunpowder) increased generalization but did not lead to a full concept formation. Our study also demonstrates how a step-by-step protocol can be used to obtain this, rather than randomly using several variations of target odor training as suggested earlier (Fischer-Tenhagen et al., 2011; Oxley and Waggoner, 2009).

Moreover, this study confirms that concept formation is not complete if the number of target odor variations is too limited. Understanding how much variation in training aids should be offered to a dog to ensure generalization to the complete range of target odors is a critical aspect of detection dog training. This study illustrates how this can be assessed in a systematic manner: by training a dog with one odor variation at a time, and by subjecting it to systematic testing after each step. In the end, detection is not done in a carousel setup but out in the field, where many other factors (e.g., environmental) influence the capability of a dog to detect the target odor. By using target odors that the dogs can recognize in a carousel setup, one at least knows that in case of a failure in the field, this is not caused by incomplete concept formation of the target odor. This requires further follow-up work.

Concept formation while retaining specificity should also be incorporated into the certification and quality control of detection dogs. As detection dogs likely encounter more target odor variations in the field then will ever be used in their training, it is important to assess whether the dogs respond to variations they have never encountered before. New and different training aids should therefore be used in certification and quality control trials to understand the possibilities and limitations of the detection dogs.

Regarding wildlife detection dogs, this study confirms that it is possible to train a dog using only captive scat to a level where it recognizes and responds to wild scat samples, even if the diets differ. This may simplify wildlife detection dog training for situations where wild scat samples are hard to obtain, although it needs to be confirmed in a controlled setting for scat of other species. In line with the quest for high training standards expressed by other authors (Johnen et al., 2013), the systematic approach presented here can be used for that purpose. Assessing concept formation during certification of a wildlife detection dog ensures quality control, which will strengthen the confidence in the use of wildlife detection dogs as a tool in ecological research.

Acknowledgments

Cobra, the dog, took center stage. Cor Oldenburg Sr. is gratefully thanked for his guidance and support during training and testing. Hans Rosenberg enabled spraint collection from captive otters in Natuurpark Lelystad and Ronald Messemaker provided spraints from wild otters in National Park Weerribben-Wieden. Using otter feces was Mariska Snelleman's idea. Freek Niewold provided expertise on otters and spraints. Anonymous reviewers helped

Table 1
Confusion matrix of the dog's responses during the criterion trials (n = 18), involving 82 samples (26 samples were not smelled by the dog)

Alert	Target odor present		Total
	Yes	No	
Yes	13	0	13
No	0	69	69
Total	13	69	82

Fresh captive spraint was the target during the first set of criterion trials. One-week-old captive spraint was the target during the second set of criterion trials.

Table 2
Confusion matrix of the dog's responses during test 3

Alert	Target odor present		Total
	Yes	No	
Yes	14	6[a]	20
No	0	137[b]	137
Total	14	143	157

The dog responded to 157 samples in 36 trials (59 samples were not smelled by the dog). In each set of 12 trials, every target odor variation was used once, resulting in fresh captive spraint, 1-week-old captive spraint, 2-week-old captive spraint, fresh wild spraint, 1-week-old wild spraint, and 2-week-old wild spraint each being used 3 times during test 3.

[a] False positives were triggered by rabbit feces (3) and hare feces (3).

[b] True negatives were performed for cattle feces (26), horse feces (21), rabbit feces (22), for hare feces (24), fox feces (19), and for roe deer feces (25).

Table 3
The number of true positive responses of the dog per target odor variation during test 3, which consisted of 3 series; every target odor was used once in every series

Target odor variation	Number of true positives
Captive, fresh	3
Captive, age 1 week	3
Captive, age 2 weeks	3
Wild, fresh	2[a]
Wild, age 1 week	1[b]
Wild, age 2 weeks	2[a]

[a] In 1 case, the trial was ended before the dog could smell this target odor due to a false positive on a sample earlier in the sequence.

[b] In 2 cases, the trial was ended before the dog could smell this target odor due to a false positive on a sample earlier in the sequence.

improve this article's concept. The idea for the article was conceived by Adee Schoon and Ignas Heitkönig. The study was designed by Adee Schoon, Cor Oldenburg Jr., and Ignas Heitkönig and carried out by Cor Oldenburg Jr., who also analyzed the data. Cor Oldenburg Jr., Adee Schoon, and Ignas Heitkönig wrote the article.

Ethical considerations

Ethical approval was not required.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Akay, M.A., 2009. Support vector machines combined with feature selection for breast cancer diagnosis. Expert Syst. Appl. 36, 3240–3247.
Browne, C.M., Stafford, K.J., Fordham, R.A., 2015. The detection and identification of tuatara and gecko scent by dogs. J. Vet. Behav: Clin. Appl. Res. 10, 496–503.
Cablk, M.E., Heaton, J.S., 2006. Accuracy and reliability of dogs surveying for desert tortoise (*Gopherus agassizii*). Ecol. Appl. 16, 1926–1935.
de Oliveira, M.L., Norris, D., Ramírez, J.F.M., Peres, P.H.D., Galetti, M., Duarte, J.M.B., 2012. Dogs can detect scat samples more efficiently than humans: an experiment in a continuous Atlantic Forest remnant. Zoologia 29, 183–186.
Fischer-Tenhagen, C., Wetterholm, L., Tenhagen, B.A., Heuwieser, W., 2011. Training dogs on a scent platform for oestrus detection in cows. Appl. Anim. Behav. Sci. 131, 63–70.
Fjellanger, R., Andersen, E.K., McLEan, I.G., 2002. A training program for filter-search mine detection dogs. Int. J. Comp. Psychol. 15, 278–287.
Ghirlanda, S., Enquist, M., 2003. A century of generalization. Anim. Behav. 66, 15–36.
Johnen, D., Heuwieser, W., Fischer-Tenhagen, C., 2013. Canine scent detection—factor fiction? Appl. Anim. Behav. Sci. 148, 201–208.
Lazarowski, L., Dorman, D.C., 2014. Explosives detection by military working dogs: olfactory generalization from components to mixtures. Appl. Anim. Behav. Sci. 151, 84–93.
Leonards, P.E.G., Zierikzee, Y., Brinkman, U.A.T., Cofino, W.P., Straalen van, N.M., Hattum van, B., 1997. The selective dietary accumulation of planar polychlorinated biphenyls in the otter (*Lutra lutra*). Environ. Toxicol. Chem. 16, 1807–1815.
Oxley, J.C., Waggoner, L.P., 2009. Detection of explosives by dogs. In: Marshall, M., Oxley, J.C. (Eds.), Aspects of Explosive Detection. Elsevier, Amsterdam, pp. 27–40.
Reber, A.S., Reber, E., Allen, R., 2009. The Penguin Dictionary of Psychology, 4th edition. Penguin Books, New York.
Schoon, A., Fjellanger, R., Kjeldsen, M., Goss, K., 2014. Using dogs to detect hidden corrosion. Appl. Anim. Behav. Sci. 153, 43–52.
Schoon, A., Haak, R., 2002. K9 Suspect Discrimination. Detselig Enterprises, Calgary.
Wasser, S.K., Hayward, L.S., Hartman, J., Booth, R.K., Broms, K., Berg, J., Seely, E., Lewis, L., Smith, H., 2012. Using detection dogs to conduct simultaneous surveys of northern spotted (*Strix occidentalis caurina*) and barred owls (*Strix varia*). PLoS One 7, 1–8.