Using Sniffing Behavior to Differentiate True Negative from False Negative Responses in Trained Scent-Detection Dogs

Astrid Concha¹, Daniel S. Mills¹, Alexandre Feugier², Helen Zulch¹, Claire Guest³, Rob Harris³ and Thomas W. Pike¹

¹School of Life Sciences, University of Lincoln, Joseph Banks Building, Lincoln LN6 7DL, UK, ²Royal Canin SAS, Avenue de la petite Camargue, Aimargues F-30470, France and ³Medical Detection Dogs, 3 Millfield, Greenway Business Park, Great Horwood, Milton Keynes MK17 0NP, UK

Correspondence to be sent to: Astrid Concha, School of Life Sciences, University of Lincoln, Joseph Banks Building, Lincoln, LN6 7DL, UK. e-mail: aconcharamirez@lincoln.ac.uk

Accepted August 3, 2014

Abstract

False negatives are recorded in every chemical detection system, but when animals are used as a scent detector, some false negatives can arise as a result of a failure in the link between detection and the trained alert response, or a failure of the handler to identify the positive alert. A false negative response can be critical in certain scenarios, such as searching for a live person or detecting explosives. In this study, we investigated whether the nature of sniffing behavior in trained detection dogs during a controlled scent-detection task differs in response to true positives, true negatives, false positives, and false negatives. A total of 200 videos of 10 working detection dogs were pseudorandomly selected and analyzed frame by frame to quantify sniffing duration and the number of sniffing episodes recorded in a Go/No-Go single scent-detection task using an eight-choice test apparatus. We found that the sniffing duration of true negatives is significantly shorter than false negatives, true positives, and false positives. Furthermore, dogs only ever performed one sniffing episode towards true negatives, but two sniffing episodes commonly occurred in the other situations. These results demonstrate how the nature of sniffing can be used to more effectively assess odor detection by dogs used as biological detection devices.

Key words: detection dogs, false negative, false positive, sniffing behavior, target odor

Introduction

Chemical detection systems are widely used to recognize the presence of a particular substance or identify low concentration of volatile compounds and hazardous gases by mimicking an animal’s sense of smell (Glatz and Bailey-Hill 2011; Oh et al. 2011; Lee et al. 2012). Although recent advances have improved the precision and efficacy of these detection technologies, these are still imperfect (Dacres et al. 2011; Zhang et al. 2013), and animals continue to appear more sensitive than man-made systems (Shelby et al. 2006; Macias et al. 2010; Weber et al. 2011; Bomers et al. 2012; Horvath et al. 2013), in addition to having the advantage of being a more dynamic system allowing quick detection over a large search area (Calbk et al. 2008). However, regardless of the nature of the detection system both false positive (where the system detects the target as present when it is absent) and false negative (where the target is present but the system fails to detect it) errors occur in these and in every detection system. The proportion of these errors occurring in a working scenario is a measure of the accuracy of the detection system. These errors may occur as a result of observation error (Mudford et al. 2009) as if the operator is not able to recognize and interpret the results obtained by the device, then the reliability of the detection system is affected. This is particularly a risk when animals are used as biological detector devices, because detection performance is assessed by handlers (Townsend 2003; Habib 2007). The dog (Canis familiaris) is the most widely employed scent-detector device for civilian and military purposes (Brook and Koehler 2003; Osterkamp 2011; Rooney et al. 2013), and these errors are well documented; both a failure to respond correctly to the presence or absence of the target odor (Bach and McLean 2003), and false negative and false positive results recorded due to handler error (Lasseter et al. 2003; Wasser et al. 2004; Lit et al. 2011). It is, therefore, important to investigate

© The Author 2014. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
factors which may help to differentiate where the error may lie. In the case of an apparent false negative, there may be a failure of the dog to detect the presence of the target odor, or a failure of the handler to recognize that the dog has detected the odor. A reduction in false negatives is particularly valuable in situations, such as when a detection dog is used for searching for a live person, detecting explosives, or identifying perpetrators of a crime.

Sniffing behavior is obviously important in the detection and discrimination of odors (Sobel et al. 2000; Verhagen et al. 2007). This is actively controlled during investigatory behavior and rapidly modulated in response to sensory input to optimize the transport of volatile compounds to the olfactory epithelium and thus for olfactory processing (Kepecs et al. 2007; Wachowiak 2011). However, whether the sniffing behavior in dogs is modified by the presence of the target odor in a scent-detection task (and so could be used as an indicator of false alerts, both negative and positive) has not been investigated. We, therefore, analyzed whether the sniffing behavior of detection dogs differs according to the olfactory detection parameters noted (true positives, true negatives, false positives, and false negatives) during a scent-detection task. It was hypothesized that when the target odor is not present, then sniffing duration will be shorter (true negatives and possibly some false positives), and that dogs will be more likely to reinvestigate marginal signals they perceive as inconclusive before issuing a response (false negatives and some false positives).

Materials and methods

Ethics statement

This research was approved by the School of Life Sciences Ethics Committee at the University of Lincoln, United Kingdom. All dogs were trained according to the ethical guidelines established by the charity Medical Detection Dogs (UK charity registration number 1124533).

Subjects

This study involved 10 detection dogs, 4 females and 6 males, ranging in age from 30 to 138 months (mean ± standard deviation [SD]: 64.3 ± 38.52 month), with body weights from 10.5 to 24.0 kg (mean ± SD: 19.24 ± 3.97 kg), and of the following breeds: Cocker Spaniel (n = 3), Labrador Retriever (n = 3), Border Collie (n = 2), and English Springer Spaniel (n = 2).

Odor sample preparation and training procedure

The dogs were trained to detect solutions of pentyl acetate (amyl acetate, CAS 628-63-7; ≥99% Sigma Aldrich, W504009) diluted in mineral oil (Sigma Aldrich, M8410) at different concentrations. A simple dilution from a stock solution of 1:1000 pentyl acetate: mineral oil (0.5 mL amyl acetate in 499.5 mL mineral oil) was used to prepare samples with concentrations above 1:100000. One to three steps of 10-fold serial dilutions of this stock solution were used to maximize the consistency of preparation of the target odor concentrations below 1:100000. One milliliter of the target concentration was required for each session and placed in a sterile 60 mL screw-top polypropylene container (4 cm diameter, item number 360133PP; Wheaton). Seven controls, each made up of 1 mL of mineral oil, were deposited in identical sterile containers. Each set of containers were used only in one session and subsequently discarded. The target and control odor containers were opened and set up on an eight-choice carousel, similar to the circular stainless steel odor presentation system which has been used in other studies (Fjellanger et al. 2002; Sargisson and McLean 2010).

Three concentrations of pentyl acetate were presented daily for each dog in a training session. The target concentrations were presented to the dogs in a systematic lowering of concentration. The rate of decrease in concentrations was 50% below the level detected earlier by the dog, based on its individual proportion of true positives obtained by concentration. During the detection training, the dogs were exposed to a range of concentrations, determined by each dog’s ability from 1:10000 to 1:1 500 000 000.

The dogs worked in an indoor room (~20 °C and 51% humidity) at the charity Medical Detection Dogs. They worked with the same handler (R.H.) throughout, and had been trained using the technique of forward chaining with a clicker and a food reward (Educ Royal Canin®). Dogs were paired on the basis of their performance in detecting similar concentrations, and each pair worked the same set of samples (target odor and controls). The order in which dogs worked (first or second) was counterbalanced during each session over different target concentrations. Sessions involved runs and passes. A run began when the target changed its position on the carousel (e.g., changed from arm 3 to 8), whereas a pass was when the dog searched the individual carousel arms 1 to 8.

The position of the target in the carousel was determined randomly for each run using custom-made computer software, and the handler was blind regarding the position of the target in the carousel and the target concentration tested. The target and controls were placed on the carousel by the same researcher (A.C.), whereas the dog and handler were in a separate room. The time between the placement of the target and controls in the carousel and the beginning of the search was between 5 and 10 min, giving time for the odors to stabilize in the headspace of their containers.

The handler and the dog entered the room together and left the room between each run, but remained inside between passes. The experimenter left the room when the handler entered. Once inside the room, the handler stood behind a screen (with a one way mirror window at a height which made
it possible for the handler to observe the dog without being seen by it) and the dog was positioned next to the handler (Figure 1). Each session consisted of two runs per concentration and two passes per run. However, a third pass was allowed when the dog did not appear to search for the position of a target on the earlier two passes. The dog could start every pass from an initial position (next to the handler) or carry on searching for a consecutive second pass. The handler, who remained behind the screen, gave a verbal command to the dog to start the search. Dogs sniffed the individual carousel arms circling either clockwise or counterclockwise without the assistance of the handler who remained behind the screen. When a dog showed the trained alert response (“sit”) at a position on the carousel, the handler confirmed the position through the computer program; if the indication of the dog was correct (true positive) it was clicked, the dog left the carousel position and returned to the initial position (next to the handler) to be rewarded by the handler with three pieces of Educ Royal Canin®. If a false positive response was given, the behavior of the dogs was not reinforced.

A new clean set of arms was placed on the carousel in each session. In addition, the carousel was cleaned with distilled water, the arms boiled, and the test room vacuumed every day to decrease the possibility of contamination, in accordance with the normal procedure of the charity.

Data analysis

The olfactory detection performance of the dog was defined in accordance with Signal-detection theory (Fjellanger et al. 2002; Macmillan and Creelman 2005; Furton et al. 2010) as follows: 1) True positive: the dog indicates the target odor in the manner in which it was trained (“sit” response), 2) False positive: the dog alerts to a nontarget position (control), 3) False negative: the dog fails to exhibit the trained alert in the presence of the target odor, and 4) True negative: the dog does not alert in the absence of the target odor.

The dogs were videotaped during every training session via a ceiling-mounted video camera (Sentient Wired® outdoor camera model N94FY), and small individual cameras (8 Channel DVR, RF2421 8ch H.263 Model 2005XA B/W EXVIEW 3.6mm CCIR [Pal]) fixed on each carousel arm (Figure 1). A total of 200 videos were pseudorandomly selected from the videos available of the detection training (Excel® random number generation), such that 20 videos were chosen for each dog including five of each of the four response types, and including a range of target concentration, from 1:700000 to 1:1500000 (pentyl acetate:mineral oil). Frames from the selected videos (with a frame rate of 25 fps) were converted to individual JPEG images using Free Studio 3 (version 5.0.28), and used to quantify both sniffing duration (s) and the number of sniffing episodes (sniffs over the odor sample). The onset of a sniffing episode was defined from when the dog’s nose was put over the hole of the carousel arm, and the end point was when the dog’s nose moved away from it.

The dog’s response type was confirmed by assessing agreement with a blind independent rater for both sniffing duration and number sniffing episodes. There was significant level of interobserver agreement between the two raters for sniffing duration for both the first ($r = 0.721, n = 20, P < 0.001$) and second episodes ($r = 0.923, n = 20, P < 0.05$).

Statistical analysis

All analyses were conducted in R.2.15.2 (http://www.r-project.org/). To determine whether sniffing duration before
a choice is made differed as a function of response choice (true positive, true negative, false positive, and false negative), we used a general linear mixed model (implemented using the lmer function of the lme4 package; Pinheiro and Bates 2000) with dog identity as a random effect. We log10-transformed the duration of sniffing data prior to analysis to ensure normally distributed residuals in the model. Tukey’s honest significant differences test (using the gln function of the multcomp package) was used to compare between levels of response choice. Differences between response choices in the number of sniffing episodes were tested using a generalized linear mixed model (using the glmer function of the lme4 package) with a binomial error distribution and dog identity as a random effect. For those sequences in which dogs performed two sniffing episodes, the difference in the log10-transformed duration of sniffing between the first and second episodes was tested using a general linear model with dog identity and episode as random effects. In all cases, statistical significance was determined by comparing full models to models lacking the independent variable using likelihood ratio tests (Crawley 2005). Pearson’s correlations were used to determine interobserver agreement between two independent raters for measuring of sniffing duration and counting sniffing episodes (Multon 2010). Results were considered statistically significant if $P < 0.05$.

## Results

The sniffing duration of dogs during the scent-detection task differed significantly between the four olfactory response choices ($F_{3,196} = 13.89, P < 0.001$). In particular, the dogs spent significantly less time sniffing true negative samples in the first episode than true positives, false positives, and false negatives (all Tukey-corrected $P < 0.001$, Table 1). Similarly the sniffing duration of false negatives was significantly shorter than true positives in the first episode (Tukey-corrected $P < 0.05$).

The presence of a second sniffing episode was observed during false positives, true positives, and false negatives, but not during true negative samples ($\chi^2_{3} = 82.79, P < 0.001$). Overall, the mean sniffing duration of the first sniffing episode across the olfactory response choices was significantly longer than the second episode ($F_{1,112} = 30.31, P < 0.001$, Table 1).

### Table 1

<table>
<thead>
<tr>
<th>Olfactory detection parameter</th>
<th>First sniffing episode (s) (mean ± SD)</th>
<th>Second sniffing episode (s) (mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positive</td>
<td>0.498 ± 0.239$^{b,c}$</td>
<td>0.257 ± 0.129</td>
</tr>
<tr>
<td>True negative</td>
<td>0.268 ± 0.118$^{a}$</td>
<td>—</td>
</tr>
<tr>
<td>False positives</td>
<td>0.468 ± 0.223$^{b}$</td>
<td>0.288 ± 0.175</td>
</tr>
<tr>
<td>False negative</td>
<td>0.408 ± 0.714$^{b,d}$</td>
<td>0.224 ± 0.096</td>
</tr>
</tbody>
</table>

Olfactory parameters with different superscript letter differ significantly from one another ($^{a,b}P < 0.001; ^{c,d}P < 0.05$) during the first sniffing episode. $N = 50$ for each parameter.

## Discussion

In detection dogs, the accuracy of the detection depends on both the dog’s olfactory capability to identify the target odor and the interpretation of the dog’s behavior by a handler. Earlier studies in detection dogs have not directly analyzed the relationship between sniffing behavior and accuracy of odor discrimination in detection tasks, concentrating instead on the total duration of the search (Thesen et al. 1993; Jeziernski et al. 2008). Our results indicate that sniffing behavior can be used alongside the trained alert response to more effectively assess detection. Specifically, we found that the sniffing duration of detection dogs used in this study is shortest when the target odor is not present and the dogs indicate this by not offering an alert response (true negative), and the dogs only ever performed one sniffing episode towards these samples compared with the other responses (true positive, false positive, and false negative). In particular, samples that resulted in true positive, false positive, and false negative decisions were sniffed for approximately twice the amount of time of true negatives. In other words, the detection dogs used in this study sniffed for twice the amount of time when the target odor was present or when it was indicated as present on a negative sample.

The shorter sniffing duration shown during true negative responses indicates that initial encoding of the presence–absence of a stimulus is rapid (Wesson et al. 2009) with discrimination determined with a single sniff (Mainland and Sobel 2006). This is comparable to observations in rodents where the sniffing duration for odor discrimination lasted between 0.15 and 0.20 s in a similar detection task (Uchida and Mainen 2003; Abraham et al. 2004; Kepes et al. 2007). Wesson et al. (2008) demonstrated that the time between the first sniff and the olfactory receptor input reaching the olfactory bulb is 0.1–0.15 s, leaving only 0.05–0.1 s for the central processing and instigation of the discriminative behavioral response. Prolonged sniffing does not seem to be necessary for the detection when the target odor is absent (true negative indications). Similar results have been described by Slotnick (2007) in rodents where longer sniffing duration was evident when determination that the target odor was present occurred. This has been interpreted as indicating that the cognitive processing for detecting whether the target odor is present or not occurs separately from the identification and...
recognition of the target odor. False positive responses can arise from the identification of background compounds similar to the target odor (Kurz et al. 1996) and the presence of extraneous odors (Bach and McLean 2003). Thus, the longer sniffing duration found in our study towards true positive, false positive, and false negative responses might reflect the engagement of higher-order pathways associated with the recognition of the odor itself.

The analysis of sniffing behavior frame by frame has been used earlier to evaluate nostril laterality in untrained dogs during the investigation of cotton swabs impregnated with different odorants (Siniscalchi et al. 2011). This technique allows a more detailed evaluation of some of the characteristics of sniffing behavior. The high level of agreement between two independent observers using this approach also shows that it is highly reliable method for objectively quantifying behavioral occurrences in extremely short periods of time. However, the application of this method for measuring sniffing duration simultaneously with the dog searching for the target odor is perhaps more limited.

Overall, the findings from this study provide evidence that sniffing behavior can be used to effectively assess olfactory alert performance in detection dogs beyond the trained alert response and was particularly valuable in differentiating true from false negative responses: an area where the consequences of error may be serious in real search scenarios such as mine and explosive detection or the search for a live person. Other aspects of dogs’ behavior regarding olfactory detection and the alert response should be investigated to identify and standardize parameters to assess dogs’ alert responses regardless of the target odor or the working situation. Future work is ongoing to further investigate the generality of the findings reported here and develop technology to evaluate sniffing behavior in real time during search tasks under field conditions.

**Funding**

This work was supported by Royal Canin SAS (grant name Nutritional factors and olfactory performance in dog).

**Acknowledgements**

The authors thank Pam Jones, Maria Jeskins-Powell, Betty Horwood, Mr Gordon, Becky Harris, and the volunteers of the charity Medical Detection Dogs for their support.

**Conflict of interest**

Royal Canin SAS supported the study financially; they approved the experimental design, financially supported Astrid Concha’s PhD candidature, and provided the food and Educan® food treats for the dogs used in this research. They did not have any influence over the study conception, design or interpretation, or the decision to publish the data. A.F. is employed by Royal Canin SAS. The authors neither have patent or stock ownership which would affect this research or publication nor have any membership of a company board of directors, membership of an advisory board or committee for a company, or consultancy for receipt of speaker’s fees from a company.

**References**


