The semiotic canine: scent processing dogs as research assistants in biomedical and environmental research

Simon Gadbois*, Catherine Reeve

Canid Behaviour Research Team, Department of Psychology and Neuroscience, Dalhousie University, Halifax - Nova Scotia, Canada

Abstract: The use of dogs in biomedical diagnosis, detection and alert as well as for the search and monitoring of species-at-risk is an emerging field of research. Standard practices are converging towards models that are not necessarily consistent with the well established field of (animal) psychophysics. We briefly discuss the different challenges of applied canine olfactory processing and discuss the adoption of more valid and reliable methods. For mostly historical reasons it seems, scent processing dogs are trained and tested using multiple alternative stimuli in choice tasks (e.g., line-ups including 6 alternative choices, or 6AFC). Data from psychophysics suggest that those methods will reduce or at the very least misrepresent the accuracy of canines. Unless canines are an exception to the rule, sensory, perceptual and cognitive arguments (e.g., Gadbois & Reeve, 2014) can be made against most multiple alternative forced choice tasks (mAFC's) in favor of detection tasks (yes/no and go/no-go procedures) or, as a compromise, simpler discrimination tasks (2AFC or 3AFC at most). We encourage the use of Signal Detection Theory as it focusses on two important factors in defining the validity and reliability of scent processing dogs: 1) It is a robust measure of sensitivity, an important factor in both diagnosis and sensory detection, and, 2) It describes the type of errors (false alarms vs. misses) that a given dog is most likely to commit, allowing for a solid assessment of performance and potentially a readjustment in training. We give an example with Diabetes Alert Dogs (DAD's) specialized in Hypoglycemia Detection in vitro and discuss the potential advantages of keeping a low number of alternatives during training and testing, the importance of low saliency training (LST), as well as adopting pure detection tasks requiring a response commitment from the dogs for both "yes" and "no" responses. The value of d' (a detectability or discriminability measure) and bias measures (criterion) are discussed in the context of canine selection, performance assessment and diagnostic accuracy across applications.

Key Words: canine olfactory psychophysics; conservation canines; diabetes; hypoglycemia; low saliency training; signal detection theory.

* *Corresponding Author*: sgadbois@dal.ca

Dogs serve increasingly important roles in a variety of medical assistance and alert positions. Due to their evolutionary close relationship with humans, dogs are very sensitive to behavioral changes and social cues from humans (Miklosi & Topal, 2013). This fact, combined with proper training, results in dogs that can accurately predict seizures (Brown & Strong, 2011), potentially predict migraines (Dawn & Bhowmick, 2013), and serve as anxiety and Post Traumatic Stress Disorder service dogs (Yount et al., 2013). Recently, researchers have become interested in whether dogs can further assist humans by using their noses to diagnose disease and alert to dangerous medical events. It is well known that dogs have incredibly sensitive noses, and empirical studies have revealed dogs detecting cancers with high levels of sensitivity and specificity (Jezierski et al., 2015). Furthermore, the use of diabetic alert dogs to signal hypoglycemic events is becoming more common. Despite a lack of empirical studies examining how dogs detect hypoglycemia, they appear to be benefitting their owners greatly (for reviews see Gadbois & Reeve, 2014; Wells, 2012). Dogs have also been involved in wildlife conservation research to monitor species-at-risk.

Dog Behavior, 3-2016, pp. 26-32
doi 10.4454/db.v2i3.43

Received, 10/31/2016
Accepted, 11/10/2016

The field of canine biomedical detection, diagnosis and alert is expanding rapidly. As teams scramble to develop methodologies, a standardization is still lacking (as discussed by Elliker et al., 2014; Jezierski et al., 2015) despite the likelihood that National Health organizations such as the FDA or Health Canada in North America will require strong Standard Operating Procedures and standardized protocols if canines are to be ever accredited as "diagnostic tools". We believe that the choice of training method will depend on the ultimate goal of the task. Currently, in the literature, the same procedures are commonly used during training conditions, testing conditions (to assess the performance of the dog), and actual diagnostic testing and field deployment. In this paper we identify a few problems in the rationale used with the most popular methods. Let us explore each issue one at a time.

The importance of understanding errors and biases

The stakes are not the same for a mine detection dog (that really cannot afford "misses" or it will pay with its life) and an endangered species search dog (missing a hidden snake during a survey search is unlikely to have dire consequences for the survival of the species or even the local population). Biomedical canines are somewhere in the middle of this spectrum: Detection dogs could be trained to find dangerous bacteria in hospital environments (Bomers et al., 2012), alert to a nocturnal hypoglycemic event with a child that cannot wear a continuous blood glucose monitor (Chen et al., 2000; Wells et al., 2008), or diagnose potential cancers (Jezierski et al., 2015). Those three functions (detect, alert, diagnose) come with different outcomes and corresponding risk assessments, in particular, the cost of making mistakes. The next sections explain what is at stake, how to measure errors and bias, and how to remediate that situation if possible and appropriate.

Not unlike decision theory and diagnostic theory, Signal Detection Theory (SDT) takes into consideration the errors made during judgements. It computes hits (true positives), correct rejections (true negatives) as well as two error types: false alarms (false positives, analogous to type 1 errors in statistics) and misses (false negatives; analogous to type 2 errors in statistics). Most diagnostic toolsets would, from these values, extract sensitivity and specificity scores. SDT goes further: It defines a very robust sensitivity index, d' ("d prime") that can be defined as an index of detectability (in detection tasks) or as an index of discriminability (in a discrimination task). This important distinction calls for a precision: Gadbois & Reeve (2014) distinguish between four psychophysical experimental contexts. We will focus here on the first three, the most likely to be used with scent dog training and assessment. The definitions given below may be slightly oversimplified in the eyes of an animal psychophysicist or sensory neuroscientist, but they cover the essentials of the current trends including one procedural option that we are suggesting. We suggest Kingdom and Prins (2016) or McNicol (2005) for a clear and concise discussion of SDT's parameters. More advanced users of SDT may want to consult Macmillan & Creelman (2005).

We will start with the most cognitive task. Most textbooks (Kingdom & Prins, 2016; Macmillan & Creelman, 2005; McNicol, 2005) discuss identification or recognition tasks. They are typically labelled "matching-to-sample" MTS tasks in the animal literature, and more specifically (implicitly at least) referring to simultaneous MTS (DMTS or delayed matchingto-sample tasks are typically used to specifically study short term memory mechanisms). Technically this task requires the handler to present a sample (standard, sometimes called a "reminder" in cognitive psychology) to the dog (that it will sniff) and then ask the dog to find the match among a number of options, typically 6 in most line-ups, although some will include 8 or 10 choices. Forensic canines (Schoon & Haak, 2009) are the typical example of this approach. As we argued in Gadbois & Reeve (2014), there are issues with this method when the intention is to determine a dog's accuracy. For example, line-ups (6+ choices) add unnecessary perceptual and mnemonic interference (see below for a discussion of interference in multiple choice tasks). In these tasks, every time the dog is asked to match the standard to one of the choices, the standard (sample) may be different. Sample sets can be very small (even just one odor presented as a cursory reminder) or n>2 with no theoretical limits. Dogs can be presented with 2, 3, 4, … n choices of one target and distractors or blanks. A classic line-up of 6 choices is therefore labelled a 6AFC (6-alternative forced choice). If high performance is expected, perceptual and mnemonic interferences are significant beyond 3 choices. This is supported by classical psychophysics (Kingdom & Prins, 2016; Macmillan & Creelman, 2005; McNicol, 2005).

Likely more common is the case of a straight discrimination between multiple options. The animal is expected to identify a target stimulus from a number of distractors (or blanks in early training). Not unlike the scenario above, the dog must choose a target among multiple choices (2AFC, 3AFC, … mAFC). The difference is that typically there is only one odor to identify, and a reminder (i.e., the sample or standard) is not offered (or necessary). This model works well when a basic perceptual discrimination is desired.

The last model, and also the simplest and potentially the most elegant, is a pure detection task. The information processing assumptions are minimal in the sense that the approach identifies a sensory sensitivity (in fact the d' mentioned above). The approach here is to present the dog with one stimulus and requires a "yes" or "no" answer. For that reason, the model is called Y/N and works within the framework of a go/no-go type of response. Note that this model is the most likely to show a bias in the decision pattern from the dog. But what seems at first like a shortcoming should be considered an advantage. A detection task will allow you to most accurately identify your dogs' response biases. If you are planning to use a 2AFC (or other mAFC procedures like line-ups and carousels) the same biases that would be identified in a detection procedure are likely to emerge. In other words, in order to understand the response bias of your dog, the detection task will give you a clearer picture of the response profile and a great context for remediation considering the simplicity of the procedure.

The Y/N model measures bias and quantifies it as a "criterion". There are different criterion measures found in the literature (see Macmillan & Creelman, 2005 for details) but the basic idea is to categorize a dog along a continuum from a "conservative" to a "liberal" decision maker. Liberal dogs are more likely to give false alarms (and minimize misses) in an attempt to maximize hits. Assuming false alarms are not problematic in the applied context where the dog works (i.e., no negative consequences), this can be a great response profile. It is certainly the one preferred by landmine detection and explosive detection dog handlers. Likewise, a conservative dog will minimize false alarms at the cost of hits, and consequently increase misses.

For now, we will focus on the consequences of knowing this information. Obviously if (and only if) a dog has a bias (and most would, the question would be "how much of a bias" or deviation from what is called in SDT an "ideal observer"), then a trainer can decide if a dog's response profile needs to be modified to change the bias. Modifications can be made by giving feedback on wrong responses, or by changing the reward saliency, frequency, or schedules.

The potential problem with proportion or percentage correct data as performance

Although SDT applies very well to Y/N and go/no-go detection tasks, it can also be applied to 2AFC and mAFC tasks. It is important to realize that some basic assumptions need to be clarified first, namely, if the responding is biased or unbiased. If the responding is unbiased, proportion correct answers are appropriate and can be transformed into a d' (see Kingdom & Prins, 2016 or Macmillan & Creelman, 2005 for computational details). The problem with assuming an unbiased response profile is that it is likely not realistic. If bias occurs, then using proportion correct as a measure of psychophysical accuracy "becomes an invalid measure of sensitivity" (Kingdom & Prins, 2016, page 161). As mentioned above, an even more fundamental issue arises: although

2AFC tasks are typically easier than Y/N tasks, mAFC tasks tend to be more challenging, with the potential exception of 3AFC tasks (Gadbois & Reeve, 2014; Macmillan & Creelman, 2005; McNicol, 2005). This can be explained fairly easily by pointing out that both sensory-perceptual and mnemonic (working memory) interference can and will occur as the number of choices presented increases. The mnemonic argument was made in Gadbois & Reeve (2014), but in the case presented there, the working memory load was significant considering that a sample set of 8 stimuli was presented (to be matched to a target in a 6AFC line-up). In most cases when one odor is presented as the sample (or standard), the main interference to worry about is sensory: when dogs sample each odor station in a line-up or carousel, there is a possible sensory interference, not excluding sensory memory especially when the stimuli are of low saliency or if they are very similar (suggesting a low d' value).

Other considerations

We will quickly address a few points before presenting an example based on ongoing research in the Canid Behaviour Research Lab at Dalhousie University. First we would like to point out that SDT is a complex area of psychophysics, sensory psychology and neuroscience and many of the important points could not be expanded upon here (see above for primers and handbook references). The computational aspect of the theory is not very complicated, but requires more space than what is allotted here to cover adequately. One issue that we will mention is the existence of non-parametric models of SDT. Although there are debates about the necessity to apply non-parametric models when they seem to be the most appropriate, some authors argue that the standard SDT theory approximates well enough non-parametric data (see Pastore et al., 2003 for a discussion). Second, the core of the issue lies in the goals of the experimenters, trainers and diagnosticians. Note that when dogs are trained for alert, the most ecologically valid task (including in training) is the detection model. Fundamentally, alert dogs need to signal the presence of the target (e.g., hypoglycemia detection dogs alert to hypoglycemia) and not respond to the absence of the target. This is a typical go/no-go situation and is closer to the Y/N decision task (except that the "no" in the go/no-go task requires no response or the inhibition or a response). In other words, alert dogs do not have an array of stimuli to "compare and contrast". They simply need to alert when the target is present, and inhibit a response when it is not (although, in assessing bias and d', you may want to consider committing the dog to a "yes" response (e.g., nose pointing the target for 5 seconds) and to a "no" response (e.g., sitting back in front the stimulus station). In other cases, it is quite possible that dogs would need to discriminate between similar stimuli that co-occur temporally and spatially. Different strains of a bacteria or parasites to detect may be examples, or as we experienced with our wildlife conservation canines, many occurring species of snakes, with only one being the main target (see Gadbois & Reeve, 2014, for the snake example).

Practical example

In Dalhousie's Canid Behaviour Research Lab, we have developed a training program that allows us to train dogs with no previous sniffer training to detect and discriminate between low saliency odors; specifically, human breath samples. What follows is a brief summary of this training program, and how it was applied in our study aimed at determining whether dogs could detect hypoglycemia in vitro, using breath samples from individuals with Type 1 Diabetes.

An important point to note is that we select our dogs very carefully. We select for dogs that are highly motivated, and that have a very high working drive. As a result, our studies (and most studies of biomedical detection with dogs) test between 3 and 5 dogs. Although this may seem like a small number of dogs with which to complete an empirical study, we are not attempting to provide evidence that all dogs are capable of doing biomedical scent detection work, but rather that a few, very carefully selected dogs can be trained to be successful.

The first phase of the training program is Low Saliency Training (LST). Here, we train our dogs to detect Orange Pekoe tea that has been steeped for 5 minutes, and then gradually decrease the saliency of the tea over time by steeping it for less time, and by diluting it with water. Using a 3AFC procedure (with a reminder), a tea stimulus is presented with two other water stimuli that serve as controls, and the dogs are required to indicate which sample is the tea sample.

If a dog demonstrates the ability to detect the tea stimulus consistently and reliably, the saliency of the tea stimulus is decreased gradually over a series of predetermined saliency levels. Once training with the liquid tea stimuli is completed, we then bridge the gap between tea and breath samples by holding tea in our mouths for 30 seconds, spitting it out, and then breathing through a breath collection tube containing a cotton ball; thus creating a "tea breath" sample. Breath samples are presented against blank cotton ball controls. Once a dog demonstrates the ability to detect the tea breath sample, they are then presented with a clean breath sample. If a dog can detect a clean breath sample successfully, the LST is complete and the dog can now detect human breath.

We find the LST phase important for two reasons: 1. It counters any potential familiarity effects by teaching the dogs to pay attention to stimuli that they have likely ignored most of their lives (human breath), and 2. The LST training serves as an inclusion test by showing us whether a particular dog is capable of detecting low saliency stimuli. If a dog cannot complete the LST successfully, we do not proceed with further training.

After completing the LST we then train the dogs to discriminate between multiple breath samples; first between breath samples from three different individuals, and then between three breath samples donated by one individual at three different times of the day. Again, this phase of training demonstrates to us that a dog is capable of discriminating between competing stimuli and that they are ready to be tested using specific medical samples. Four volunteer dogs, Nutella, Koda, Bella, and Mist, successfully completed this training program. When we presented the dogs with breath samples donated by individuals with Type 1 Diabetes, we first tested their ability to discriminate between three different breath samples obtained from one individual by presenting them with three breath samples simultaneously: one when the blood sugar of the breath donor was hypoglycemic, one when it was normal, and one when it was hyperglycemic, and requiring that the dog identify the hypoglycemic breath sample. We tested their ability to do this with sample sets from three different individuals. All four of the dogs tested were able to discriminate between the samples with average accuracy rates between 90% and 100%.

We then trained Nutella and Koda to detect hypoglycemia by presenting them with low, normal, and high samples from one individual sequentially (one sample at a time). Here, the dogs were trained to smell a single sample and indicate whether "yes" this is a hypoglycemic sample, or "no" this is not a hypoglycemic sample. Once they demonstrated the ability to detect the hypoglycemic sample within a sample set, we then added a second sample set (a second hypoglycemic breath sample, a second normal breath sample, and a second hyperglycemic breath sample) from the same individual, and tested whether Nutella and Koda could generalize the odor of hypoglycemia to the second sample set; that is, say "yes" to both low breath samples and "no" to all other samples. Presenting the samples in this way allowed us to test whether the dogs could identify multiple instances of hypoglycemia occurring in one individual. As illustrated by the data in Table 1, Nutella was capable of generalizing the odor of low blood sugar to the new breath sample. Koda, however, was not, as illustrated by his low sensitivity score. Although Koda continued to signal "yes" to the first low blood sugar breath sample, he never signaled "yes" to the second low blood sugar breath sample from the same individual.

Table 1. Nutella and Koda's performance on a test of their ability to generalize the odor of hypoglycemia across two breath samples from one individual, using a Go/No-Go procedure.

	Nutella	Koda
d'	1.675	1.468
C	0.313	0.911
Sensitivity	70%	43%
Specificity	88%	95%
Accuracy	78.7%	69%
Precision	84.8%	89%
No. of trials	160	120

As illustrated by the dogs' performance across the two sample presentation methods, presenting the stimuli to the dogs using both 3AFC and Y/N presentation procedures allows for a more thorough understanding of their abilities. Although both Nutella and Koda were able to discriminate between samples successfully (both averaged 100% accuracy), when tested on their ability to detect (Y/N) low blood sugar samples, only Nutella continued to be successful. The "C" in Table 1 represents the criterion (one of the few measures of bias). A positive C indicates a conservative decision maker. The values are from -1 (very liberal) to +1 (very conservative). Both dogs are conservative, but Nutella is the closest to the "Ideal Observer" (meaning that she maximized both correct rejections and hits). Note that Koda has very high specificity. This means that he is accurate at indicating what is not a hypoglycemic sample (maximizing correct rejections) but poor at identifying what is a hypoglycemic sample (he committed more misses).

Conclusion

When assessing the ability of dogs to diagnose, detect or alert, clear context-appropriate goals need to set before considering a training and assessment protocol. If the goal is to obtain an accurate diagnosis, then procedures resulting in high accuracy (Y/N) and a clear description of the errors and bias need to be adopted (SDT). Even if perfect or closeto-perfect accuracy is not essential, a procedure that can identify error types is still very informative and can influence training or help in selecting "top performance" dogs. Multiple choice procedures such as line-ups increase the sensory and mnemonic interference of the task while reducing performance (percentage correct scores) which may be appropriate during training, but fail to give an accurate profile of the performance of the dog.

References

Blough D., Blough P. 1977. Animal psychophysics. In: Honig, W.K., Staddon, J.E.R. (Eds.), Handbook of operant behaviour. Englewood Cliffs, NJ: Prentice-Hall, Inc, 514-539.

Blough D.S., Lipsitt L.P. 1971. The discriminative control of behavior. In: Kling, J.W., Riggs, L.A. (Eds.), Experimental Psychology, 3rd ed. NY: Holt, Rinehart and Winston, 743-792.

Brown S.W., Strong V. The use of seizure-alert dogs. Seizure. 2011;10: 39-41.

Bomers M.K., van Agtmael M.A., Luik H., van Veen M.C., Vandenbroucke-Grauls C.M.J.E., Smulders Y.M. Using a dog's superior olfactory sensitivity to identify Clostridium difficile in stools and patients: proof of principle study. The B.M.J. 2012; 345: 1-8.

Chen M., Daly M., Williams N., Williams S., Williams C., Williams G. Non-invasive detection of hypoglycaemia using a novel, fully biocompatible and patient friendly alarm system. The B.M.J. 2000; 321: 23-30.
Dawn M.A., Bhowmick A. Survey of migraine sufferers with dogs to evaluate for canine migraine-alerting behaviors. The Journal of Alternative and Complementary Medicine. 2013; 19: 501-508.
Elliker K.R., Sommerville B.A., Broom D.M., Neal D.E., Armstrong S. Williams H.C. Key considerations for the experimental training and evaluation of cancer odour detection dogs: lessons learnt from a double-blind, controlled trial of prostate cancer detection. B.M.C. Urology. 2014; 14: 1-9.
Engen T. Psychophysics 1: Discrimination and detection. 1971; 11-46.
Gadbois S., Reeve C., 2014. Canine olfaction: Scent, sign, and situation. In: Horowitz, A. (Ed.), Domestic dog cognition and behavior. Springer, New York, 3-29.
Jezierski T., Walczak M., Ligor T., Rudnicka J., Buszewski L. Study of the art: Canine olfaction used for cancer detection on the basis of breath odour. Perspectives and limitations. Journal of Breath Research. 2015; 9: 1-12.
Kingdom F.A.A., Prins N., 2016. Psychophysics. A practical introduction. Elsevier.
Macmillan N.A., Creelman C.D., 2005. Detection theory. A user's guide (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
McNicol D., 2005. A primer of signal detection theory. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Miklósii A., Topàl J. What does it take to become 'best friends'? Evolutionary changes in canine social competence. Trends in Cognitive Sciences. 2013; 17: 287-294.
Pastore R.E., Crawley E.J., Berens M.S., Skelley M.A. "Nonparametric" A' and other modern misconceptions about signal detection theory. Psychon. Bull. Rev. 2003; 10: 556-569.
Schoon G.A., Haak R., 2002. K9 suspect discrimination: Training and practicing scent identification line-ups. Calgary, Alberta: Detselig Enterprises.
Wells D. Dogs as a diagnostic tool for ill health in humans. Alternative Therapies in Health and Medicine. 2012; 18: 12-17.
Wells D., Lawson S.W., Siriwardena N. Canine responses to hypoglycemia in patients with Type 1 Diabetes. The Journal of Alternative and Complementary Medicine. 2008; 14: 1235-1241.
Yount R., Ritchie E.C., St. Laurent M., Chumley P., Olmert M.D. The role of service dog training in the treatment of combat-related PTSD. Psychiatric Annals. 2013; 43: 292-295.

La semiotica canina: i cani per la ricerca olfattiva come assistenti nella ricerca biomedica ed ambientale

Simon Gadbois*, Catherine Reeve

Canid Behaviour Research Team, Dipartimento di Psicologia e Neuroscienze, Università di Dalhousie, Halifax, Nuova Scozia - Canada

Sintesi

L'utilizzo dei cani nelle diagnosi cliniche e nella ricerca e monitoraggio di specie a rischio, è un campo emergente di ricerca.

Le procedure standard convergono verso modelli che non sono necessariamente in sintonia con il campo della psicofisica animale. In questa review saranno descritti i problemi relative alla ricerca olfattiva e l'adozione di metodi validi.

Per ragioni storiche, i cani sono addestrati utilizzando stimoli multipli in test di scelta (ad esempio stimoli in fila che includono 6 scelte alternative, 6AFC).

I risultati di ricerche psicofisiche suggeriscono che questi metodi riducono l'accuratezza del cane. a meno che i cani non costituiscano un'eccezione alla regola, prove sensoriali, percettive e cognitive possono essere addotte contro i test di scelta multipli (mAFC) in favore di prove di rilevamento (sì/no) o prove di discriminazione più semplici (2-3 stimoli).

Gli autori incoraggiano ad usare la Teoria di rilevamento dei segnali poichè si basa su due importanti fattori nel definire la validità e affidabilità del cane da ricerca olfattiva:

1) è una misura robusta della sensibilità, un fattore importante nella diagnosi e nel rilevamento sensoriale;
2) descrive i tipi di errore (falsi allarmi vs mancate segnalazioni) che un cane può commettere, permettendo un'accurata valutazione della performance e potenzialmente una modificazione del percorso di addestramento.

Sarà fornito un esempio di ciò descrivendo l'addestramento dei cani per l'allerta diabete (DAD) specializzati nel rilevamento dell'ipoglicemia. Sarà discusso il potenziale vantaggio di mantenere ridotto il numero di stimoli alternativi durante l'addestramento e di un training con stimoli poco salienti.