Behavioural Processes xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

Tuberculosis detection by pouched rats: Opportunities for reinforcement under low-prevalence conditions

Timothy L. Edwards[a,*], Haylee Ellis[a,b], Erin E. Watkins[c,b], Christiaan Mulder[b,d], Georgies Mgode[b,e], Christophe Cox[b], Alan Poling[c,b]

[a] University of Waikato, School of Psychology, Private Bag 3105, Hamilton 3240, New Zealand
[b] Anti-Persoonsmijnen Ontmijnende Product Ontwikkeling (APOPO), PO Box 3078, Morogoro, Tanzania
[c] Western Michigan University, Department of Psychology, 1903 W Michigan Ave, Kalamazoo, MI 49008, USA
[d] Amsterdam Institute for Global Health and Development, Trinity Building C, PO Box 2270, 1100 DE Amsterdam, The Netherlands
[e] Sokoine University of Agriculture, Pest Management Centre, PO Box 3110, Morogoro, Tanzania

ARTICLE INFO

Keywords:
Concept formation
Cricetomys ansorgei
Discrimination learning
Pulmonary tuberculosis
Scent detection
Techniques and procedures
Diagnostic

ABSTRACT

Giant African pouched rats (*Cricetomys ansorgei*) have been employed successfully in two operational tuberculosis-detection projects in which they sniff sputum samples from symptomatic individuals who have visited tuberculosis clinics. The prevalence of pulmonary tuberculosis in this population is high, approximately 20% in the regions where the rats have been used. If the rats are to be used to screen individuals from lower-prevalence populations, their performance under such conditions must first be evaluated. In this study, the prevalence of tuberculosis-positive samples presented to eight pouched rats was reduced to approximately 5%, and the percentage of known-positive samples included as opportunities for reinforcement was varied in sequence from 10 to 8, 6, 4, 2, 4, and 2. Liquid food reinforcers were delivered for identification responses to known-positive samples and at no other time. The rats' accuracy was clinically and statistically significantly lower at 2% than at the other values. These results indicate that the rats can perform well in low-prevalence scenarios but, if they are used under the conditions of the present study, at least 4% of the samples presented to them must be opportunities for reinforcement.

1. Introduction

Giant African pouched rats have been successfully trained and deployed as detectors of human tuberculosis (TB) in sub-Saharan Africa (Edwards, Valverde, Mulder, Cox, & Poling, 2016; Mahoney et al., 2011; Poling et al., 2010, 2011; Weetjens et al., 2009). The rats have screened sputum samples that were previously evaluated at TB clinics in Tanzania since 2007 and in Mozambique since 2013.

The individuals who visit TB clinics generally present with TB symptoms, and in our previous studies approximately 20% were confirmed as TB-positive (e.g., Mahoney et al., 2011; Poling et al., 2010). Sputum smear microscopy is the standard diagnostic tool employed by TB clinics in the developing world, but the sensitivity of this technology as it is applied in clinical settings (i.e., the proportion of TB-positive individuals identified as such) is notoriously low, often around 50% (Reid and Shah, 2009). TB-positive individuals without a clinical diagnosis do not receive treatment and continue to spread the infection to others. We recommend that readers who are unfamiliar with TB refer to the relevant fact sheet produced by the World Health Organization for supplementary information (World Health Organization, March 2017).

APOPO collects sputum samples from 28 clinics in Tanzania and 14 clinics in Mozambique at the time of this writing. Upon arrival at APOPO's TB laboratories, the samples are heat-inactivated. Sample information is entered into a database, which is used to arrange rat evaluation sessions by randomly interspersing samples that are known to be positive based on clinical microscopy results as 10% of the total number of samples. The sensitivity of clinical microscopy is low, but the positive predictive value (the proportion of microscopy-positive samples that are from TB-positive individuals) is high. Therefore, clinic-positive samples are used to create opportunities for reinforcement of correct indications. In this discrete-trials arrangement, an opportunity for reinforcement is present on a proportion of trials, those where a clinic-positive sample is presented, and absent on the remaining trials. Trainers arrange differential reinforcement by presenting food following the indicator response (nose in the hole above the sample for at least 3 s) when clinic-positive samples are present and not when they are absent. Food is not, however, presented when a rat responds to a clinic-

* Corresponding author.
E-mail address: tim.edwards@waikato.ac.nz (T.L. Edwards).

http://dx.doi.org/10.1016/j.beproc.2017.05.015
Received 25 January 2017; Received in revised form 18 April 2017; Accepted 22 May 2017

negative sample that is actually TB-positive, because trainers do not know the true status of such a sample. Therefore, correct indications are differentially but intermittently reinforced.

In an alternative conceptualization of this arrangement, clinic-positive samples are discriminative stimuli in the presence of which the indication response is differentially reinforced. Indications of clinic-negative samples are instances of stimulus generalization. Generalization to clinic-negative samples from TB-positive individuals is desirable (this is the basis of the rats' clinical value), but generalization to clinic-negative samples from TB-negative individuals is undesirable because all clinic-negative samples that are indicated as positive by the rats must be evaluated with confirmatory technology, and evaluation of samples from TB-negative individuals reduces operational efficiency. From this conceptual perspective, TB detection is much like concept formation in that there are no distinct boundaries between exemplars and non-exemplars but that, with appropriate training, better-than-chance accuracy with novel exemplars can be achieved (e.g., Herrnstein, Loveland, & Cable, 1976). Comparisons of the rats to gold standard technology (culture) suggest that stimulus generalization is sufficient to capture many but not all samples from TB-positive individuals that have been incorrectly classified at TB clinics (Mahoney et al., 2012; Reither et al., 2015). These studies also suggest that generalization to samples from TB-negative individuals is common but that accuracy with clinic-negative samples is significantly better than the degree of accuracy that would be expected by chance.

Any samples with unknown status that are indicated as positive by the rats are marked for confirmatory analysis in APOPO's laboratory, which involves sample centrifugation and evaluation using fluorescence microscopy. This method of microscopy for TB diagnosis has been found to be more sensitive than direct sputum smear microscopy as conducted in clinical settings (Steingart et al., 2006; Uddin et al., 2013). For example, Uddin et al. found fluorescence microscopy sensitivities of 70.5% and 82.9% with direct (non-centrifuged) and concentrated (centrifuged) smear techniques, respectively.

Patients with samples that are clinic-negative (i.e., TB-negative according to clinical microscopy and other diagnostic methods) but found to be positive in APOPO's laboratory (i.e., TB-positive according to concentrated smear fluorescence microscopy following positive identification by rats) are contacted and instructed to return to the clinic to initiate treatment. In 2014, these activities resulted in the discovery of 2057 TB-positive individuals who were incorrectly classified as TB-negative at clinics, and increased case detection rates in collaborating clinics by 53% in Mozambique and 39% in Tanzania (Poling et al., 2017). Given the impact of the rats in this enhanced-case-detection role, application of the TB-detection rats for other activities, such as active case finding in high-risk populations, is of interest. However, active case finding in high-risk but not necessarily symptomatic populations would require the rats to evaluate samples from populations with a lower TB prevalence than the prevalence associated with the current operational scenario in TB clinics. For instance, the estimated prevalence of TB among mineral miners in sub-Saharan Africa, one such high-risk population, is between 3% and 7% (Stuckler, Basu, McKee, & Lurie, 2011).

Additionally, under lower TB-prevalence conditions, practical considerations will undoubtedly limit the number of confirmed positive samples that will be available to serve as opportunities for reinforcement. If a procedure were available that isolated many such samples, the rats would be of no clinical value. Other methods of arranging additional opportunities for reinforcement, such as maintaining a collection of samples with known status that can be interspersed among samples with unknown status, are also problematic because different sample collection, handling, and storage procedures will result in irrelevant differences between known-status and unknown-status samples that can quickly come to control the indication response instead of relevant disease-associated stimuli.

Unfortunately, little is known about the performance of scent-detecting animals under conditions where few opportunities for reinforcement are arranged. In one relevant study, Sargisson and Mclean (2010) studied dogs in a remote explosives scent tracing task in which the relative number of positive samples was held constant at 1–6% of the total samples presented in individual sessions. They examined the effects of manipulating the percentage of identification responses (sits) to positive samples that were reinforced. High (60–75% of responses reinforced), moderate (35–50%), and low (20–30%) percentages were examined. Results indicated that the dogs missed significantly more positive stimuli (i.e., failed to respond to them) under the low-percentage condition than under the other two conditions, but false alarms (indication responses to negative stimuli) did not vary significantly across conditions. These findings suggest that scent-detecting dogs can function well when reinforcers are rarely available or delivered, but their performance deteriorates when reinforcers are too scarce. No comparable data have been obtained with pouched rats. The purpose of the present evaluation was to determine the minimum proportion of total samples that can be arranged as opportunities for reinforcement that will maintain accurate responding in rats detecting tuberculosis in a simulated low-prevalence scenario.

2. Methods

2.1. Subjects, setting, and apparatus

Five female and three male giant African pouched rats (*Cricetomys ansorgei*) between 1.5 and 7.25 years of age (Mean = 3.4, SD = 2.25) that had previously been used for operational TB detection in Morogoro, Tanzania evaluated samples in this study. The rats had *ad libitum* access to fresh water outside of experimental sessions and were housed individually or in pairs in ~1-m^3 living cages equipped with running wheels, clay nesting chambers, wooden structures for climbing and gnawing, and bedding material. To increase the reinforcing effectiveness of food provided in experimental sessions, rats were not fed prior to experimental sessions and total daily food intake was restricted. APOPO's Animal Welfare Assurance was approved by the Office of Laboratory Animal Welfare (A5720-01).

The procedures that were used for initial training of the pouched rats have been detailed elsewhere (Poling et al., 2011). In brief, young rats were habituated to human handling and a variety of stimuli, then placed in the evaluation chamber where they were trained with an auditory click and food to "indicate" by placing and holding their noses in holes in the floor above sputum samples from TB-positive individuals. Once the rats were reliably indicating single samples, additional samples, both positive and negative, were systematically introduced to the training sessions. Only indications of TB-positive samples were reinforced and, when the rats reached predetermined accuracy criteria, the number of samples presented in each "run" was increased until the rats were evaluating 10 samples at a time.

Samples were inserted into cartridges that held 10 samples each in the order specified by a database, which randomly assigned samples to positions in the sample array. The cartridges were locked in place beneath 10 holes in the floor of a 205-cm long, 55-cm wide, and 55-cm high, aluminium and acrylic glass cage.

2.2. Samples

We reduced the prevalence of positive samples in the set of samples used in the present experiment by selecting samples that were indicated as TB-negative by clinical microscopy and that had been evaluated by a separate group of 10 rats but indicated by no rats in that group (henceforth referred to as "unknown-status samples"). Based on comparison to culture data, the expected prevalence of TB-positive samples in this sample set is approximately 5% (Mahoney et al., 2012). Positive sample prevalence and the relative number of opportunities for reinforcement were adjusted by inserting a proportion of samples