Using Behavior-Analytic Techniques to Benefit an African Non-Governmental Organization: Improving and Expanding Services

Amy L. Durgin
Western Michigan University, durginal@gmail.com

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Industrial and Organizational Psychology Commons

Recommended Citation

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
USING BEHAVIOR-ANALYTIC TECHNIQUES TO BENEFIT AN AFRICAN NON-GOVERNMENTAL ORGANIZATION: IMPROVING AND EXPANDING SERVICES

by

Amy L. Durgin

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements for the
degree of Doctor of Philosophy
Department of Psychology
Western Michigan University
April 2013

Doctoral Committee:

Alan Poling, Ph.D., Chair
Ron van Houten, Ph.D.
Heather McGee, Ph.D.
Steven Ragotzy, Ph.D.
The combination of a slowly recovering global economy, increased corporate competition, and higher standards from donors with respect to governance and accountably have posed significant challenges for nonprofit organizations around the world. In order to survive, these organizations must adapt their operational models and find new strategies for delivering on strategic goals, improving operational efficiency, and differentiating their services. The purpose of the present project was to employ behavioral techniques to improve operational efficiency and to develop opportunities for organizational growth by expanding the range of services provided by a nongovernmental organization (NGO) in East Africa that uses scent-detection rats for humanitarian purposes. The project comprised three experiments. In the first experiment, a package intervention involving a job aid and feedback training was developed to improve the performance of staff in evaluating and conducting animal training sessions. Results suggested that both supervisor and staff performance improved as a result of the package intervention. Moreover, the intervention appeared to be sustainable and cost-effective. In the second and third experiments, two new applications of scent-detection rats were systematically evaluated. Results of these two studies provide proof-of-principle with respect to the rats’ ability to find people and to detect salmonella bacteria in horse feces. Although challenging, the present project was successfully completed and demonstrates
that the same general strategies used to benefit other kinds of organizations can be of value to NGOs operating in resource-poor and culturally diverse areas.
ACKNOWLEDGEMENTS

Early in my graduate career, my advisor, Al Poling, said to me, "You have to be resistant to punishment. Do that, and you'll be fine." At the time, I thought the advice just referred to dealing with manuscript rejections from journals. More importantly, however, it turned into advice that I would carry with me while working in Tanzania for two years and heed on a daily basis. I was extremely fortunate to work under Al's wing. He encouraged me to seek opportunities outside the laboratory, which led me into the school districts and eventually Africa. His brilliance, humility and generosity are of a truly exceptional kind. He has found a unique niche as a friend and inspiration to me, and I can only hope to become half the person and behavior analyst that he is. I look forward to trying, Al.

Secondly, I would like to thank my dissertation committee -not only for their patience and helpful feedback on my dissertation -but also for the various ways in which they have each helped shape my academic career. Heather McGee was a major influence in getting me involved in organizational behavior management. She skillfully helped teach me the application of concepts and tools that I have been using ever since my Behavioral Systems Analysis class with her. Ron Van Houten showed me in my first graduate course, that exhibiting passion for what you love to do can be contagious, admirable and unforgettable. Steve Ragotzy demonstrated an inspiring level of dedication towards helping students learn and become leaders in their own lives. I hope to be able to do the same. I would also like to thank Alyce Dickinson for being a wonderful teacher and a powerful source of guidance and encouragement.
Acknowledgements--Continued

To all of the staff working at APOPO, who listened when I said "Hey guys, this is an experiment! Come on!" -thank you (asanteni). I worked with an amazing team at APOPO who taught me never to take life too seriously.

I would like to thank the really good friends I had in my graduate program as well as those who worked by my side in Tanzania. From relaxing on the front porch after a long day at the "office", to handling tumultuous moments at the market in town - I could not have survived without them.

Finally, I would like to thank the people that have been supporting me unconditionally since the day I was born. To my siblings and my parents, I am very lucky to have you in my life indefinitely. To my parents, two intelligent and fun individuals who make an excellent team, thank you for everything. I owe you some money for cat food.

Amy L. Durgin
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ............................................................................................................ ii
LIST OF TABLES ..................................................................................................................... vi
LIST OF FIGURES ................................................................................................................... vii
INTRODUCTION ....................................................................................................................... 1
EXPERIMENT ONE .................................................................................................................... 8
  METHOD ............................................................................................................................... 13
  RESULTS ............................................................................................................................. 23
  DISCUSSION ......................................................................................................................... 34
EXPERIMENT TWO ................................................................................................................... 41
  METHOD ............................................................................................................................... 43
    Experiment 2a .................................................................................................................. 49
    Experiment 2b .................................................................................................................. 51
    Experiment 2c .................................................................................................................. 55
  DISCUSSION ......................................................................................................................... 57
EXPERIMENT THREE ............................................................................................................... 59
  METHOD ............................................................................................................................... 61
  RESULTS ............................................................................................................................. 68
  DISCUSSION ......................................................................................................................... 71
GENERAL DISCUSSION .......................................................................................................... 73
FUNDING .................................................................................................................................. 76
<table>
<thead>
<tr>
<th>REFERENCES</th>
<th>77</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPENDICES</td>
<td></td>
</tr>
<tr>
<td>A. Human Subjects Institutional Review Board Approval Letter</td>
<td>89</td>
</tr>
<tr>
<td>B. The Job Aid</td>
<td>91</td>
</tr>
<tr>
<td>C. Treatment Acceptability Rating Form-Revised</td>
<td>94</td>
</tr>
<tr>
<td>D. Diagram of the Semi Automated Cage</td>
<td>96</td>
</tr>
<tr>
<td>E. Photos of Training the Search Rats</td>
<td>98</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1. Training behaviors and results identified by the task analysis…………………..18
2. Percent scores for supervisors on individual target behaviors across each phase……25
3. Percent scores for trainers on individual target behaviors across each phase…………28
4. Percent scores for trainers on individual target behaviors at a secondary
   training stage prior to and post intervention……………………………………………30
LIST OF FIGURES

1. Performance scores (percent correct responses) for trainers and supervisors (summed across rats) during each session of the study. .................................................................24

2. Performance scores for all three trainers at a field training stage, where the intervention was not used, prior to and following the intervention. ............................................29

3. Percent agreement scores between supervisor and consultant........................................32

4. Responses of each supervisor to the five questions used to evaluate treatment acceptability. ................................................................................................................33

5. The semi-automated cage is shown in the top panel and a rat indicating on a sample in the lower panel.................................................................45

6. Average sensitivity and specificity data for the rats as a group in the culture broth only phase (A), the dried horse manure phase (B) and the fresh horse manure phase (C) ........................................................................50

7. Average sensitivity and specificity data for the rats as a group with Salmonella Newport (A) or Gallanarium (B) samples were presented. ...................................................53

8. Average sensitivity data for the rats as a group when presented with Newport, Gallinarum (GLT) simultaneously as well as two novel strains – Saintpaul and Agona. ................................................................................56

9. The number of days taken for each rat to reach the learning criteria in each of the four training stages ......................................................................................68

10. Percentage of trials in which rats in the experimental and control groups located human targets within 30 s ....................................................................................70

11. Time (s) elapsed from the onset of the return command to the rat's return to the release point. ..................................................................................................................70
INTRODUCTION

Today, organizations in the nonprofit sector face increasing scrutiny regarding the efficiency and effectiveness of their operations (Edwards, 2010; Pallotta, 2008). At the same time there has also been rapid growth within the nonprofit sector and shifts in government and donor funding, as well as competition from for-profit organizations offering human services (Chetkovich & Frumkin, 2003). These changes have created a very competitive funding environment for nonprofit organizations. As a result, there is significant pressure for these organizations to adapt their operational models in order to maintain a competitive edge. Two commonly proposed strategies for adapting involve partnering with for-profit corporations and adapting more business-like procedures within the nonprofit context. The strategies are not necessarily mutually exclusive, and each offers potential benefits and poses challenges.

Partnerships between nonprofits and for-profits allow greater access to management expertise and support as well as additional financial resources (Waddell, 2000). There is, however, concern that such partnerships will increase the likelihood of co-optation and compromise the autonomy of nonprofits (Baur & Schmitz, 2012). Similarly, common for-profit business models can be tailored to fit the unique needs and cultures of nonprofit organization and increase both their effectiveness and their appeal to some donors. Some donors, however, may react negatively to these models and individuals working in the nonprofit sector are often reluctant to adopt such models because they feel that doing so will conflict with or divert efforts from their social missions (Mannell, 2001; Weerawardena, McDonald & Mort, 2010).
In recent years, there has been much discussion regarding whether nonprofits are fundamentally different than for-profits (e.g., Frumkin & Andre-Clark, 2000; Hirth, 1997), why nonprofits characteristically have trouble tracking efficiency and impact (Harsh, Mbatia & Shrum, 2010; Nunnenkamp & Ohler, 2012), and the best ways to improve such organizations (Chetkovich & Frumkin, 2003; Mannell, 2010), but few relevant data have appeared. The purpose of the present project, detailed later, was to use tactics and strategies characteristic of behavior analysis and the related field of organizational behavior management (OBM) to improve and expand the services of a nongovernmental organization (NGO) located in East Africa and to empirically document the outcome of selected activities. These tactics and strategies have proven useful in improving many aspects of organizational function in a wide range of settings (Johnson, Redmon, & Mawhinney, 2001) and our research team believed that they could be readily adapted to meet the needs of the NGO.

**Background**

What was once viewed as merely the charitable, philanthropic or voluntary sector, the nonprofit sector has become increasingly influential and is rapidly growing. According to *The Nonprofit Almanac 2011* published by the Urban Institute, 1.4 million nonprofits were registered with the US Internal Revenue Service (IRS) in 2009 - an increase of 19 percent from 1999. By 2012, nonprofit organizations registered with the IRS grew to 1.6 million (Roegar, Blackwood & Pettijohn, 2012). In terms of employment and wages, nonprofit organizations grew by 4 and 6.5 percent, respectively, between 2007 and 2010, substantially surpassing the for-profit and governmental sectors in rate of growth. In 2011 the nonprofit community contributed approximately 5.5% to
the American nation's entire Gross Domestic Product (GDP), roughly equivalent to $805 billion (Roegar, Blackwood & Pettijohn, 2012).

Although some view the term synonymous with nonprofit, NGOs are typically distinguished by their efforts in providing some form of international development or humanitarian service. The World Bank (1995) defines NGOs as "private organizations that pursue activities to relieve suffering, promote the interests of the poor, protect the environment, provide basic social services, or undertake community development" (p. 13). Thus, NGOs make up a wide variety of organizations working to improve the welfare of poor communities around the world.

NGOs have existed in one form or another since the eighteenth century, and their role in abolishing the slave trade is generally acknowledged (Edwards & Hume, 1992). In the 1980s, the reputation and notoriety of NGOs reached an all-time high. As Scholte and Schnabel (2002) pointed out, “The 1980s saw the onset of what could arguably be described as the golden age of the international NGO” (p. 250). During that time, NGOs were widely recognized across political and socio-economic spectrums for being cost-effective, flexible, democratic, and in some cases, alternatives to weak or corrupt government (Clark, 1991; Harsh, Mbatia & Shrum, 2010; Korten, 1990). NGOs were esteemed for their innovative solutions to longstanding development problems, their willingness to adapt to local needs and conditions, and their ability to reach communities unlike any other institutional body (Edwards & Hulme, 1992). The number of articles and books published on NGO development increased exponentially across the decade as university programs and training centers for international studies sprang up (e.g., Global Partnership for NGO Studies, Education and Training, and the School for International
Training, USA) (see also van Rooy, 2000). Larger NGOs such as Oxfam, Greenpeace, or Doctors without Borders became publicly recognizable names. As Iriye (1999) stated, to ignore the rise and influence of NGOs is to "misread the history of the twentieth-century world" (p. 424).

NGOs remain an important and large-scale presence in international development today, however, they are currently facing serious scrutiny regarding the legitimacy and accountability of their operations (Cooley and Ron, 2002; Ebrahim, 2003; Edwards & Hume, 1995; Fisher, 1997; Najam, 1996; van Rooy, 2000; Walsh & Lenihan, 2006). The concern is due, in part, because despite the scholarly fervor of the past, there is little empirical evidence to support the previously held beliefs concerning NGOs’ stellar performance and positive impact (Lewis, 2001; Mitlin, Hickey & Bebbington, 2007; UNRISD, 2000; van Rooy, 2000). Most experts agree the supposed legitimacy of NGOs in the past was based more on "a belief in value-driven organizations than on actual monitoring and assessment of their accomplishments” (Ebrahim, 2003, p. 813).

Issues surrounding organizational legitimacy arguably stemmed from various systematic problems as well as the lack of formal requirements for NGOs to keep detailed reports of their programs, the lack of any generally acceptable or useful measure of impact, and the lack of resources to support measurement and operational reporting in the first place. Nevertheless, the international development and policy discourse quickly turned into theoretical discussions on how to improve the impact of NGO activities (Edwards & Hulme, 1992; Fowler, 1997, 2000; Lewis, 2001), how to assess and develop transparent practices (Doyle & Pattel, 2008; Ebrahim, 2003; Najam, 1996), and how to better manage NGOs’ relationships with local governments and organizations (Groves &
Hinton, 2004; Robinson, Hewitt, & Harriss, 2000). In recent years, donors have raised their expectations for NGOs and now characteristically require them to demonstrate efficient use of resources, workable corporate practices, and documented effectiveness (Edwards, 2010; Pallotta, 2008). The rapid growth in the number and diversity of NGOs, as well as the greater expectations for more transparent and effective practices, both occurring in an uncertain global economy, have created a fiercely competitive funding environment for NGOs. NGOs will need to develop cost-effective practices for demonstrating efficient resource use (human and financial) in order to create the competitive edge necessary to maintain funding and to survive.

Although there is no doubt that a great deal of work has been done to improve organizational operations within NGOs, the published work primarily comprises case studies (see Arevalo, Ljung & Sriskandarajah, 2009; Gomes & Liddle, 2009; Nair & Vohra, 2010), survey-based research (see Borwankar & Velamuri, 2009; Weerawardena, McDonald & Mort, 2010), and descriptions of theoretical frameworks for improving organizational performance (Harsh, Mbatia & Shrum, 2010). Few adequately controlled experiments have appeared and, as discussed later, studies using the tactics and strategies of behavior analysis are sorely lacking.

Purpose of the Present Project

The purpose of the present project was to employ behavioral techniques, which rely on the systematic and data-based use of learning principles to change behavior in desired ways, to improve operational efficiency and to develop opportunities for organizational growth by expanding the range of services provided by the organization. The project comprised three experiments. The first improved performance of
organizational staff engaged in an activity crucial to the primary service provided by the NGO. The other two provided proof of principle with respect to two other important services that the NGO potentially could provide. Collectively, by utilizing a systematic and data-based approach, the three studies demonstrated the worth of the organization's services to stakeholders, both donors and direct recipients of their services, and improved the efficiency with which those services were provided.

The NGO

The project was conducted under the auspices of Anti-Persoonsmijnen Ontmijnende Product Ontwikkeling (APOPO), which is Dutch for Anti-Personnel Landmines Detection Product Development. APOPO is a Belgian, NGO located in Tanzania, a resource-poor country in East Africa. The goal of the organization is to develop and disseminate scent-detection rat technology for humanitarian purposes; specifically, for detecting landmines in hazardous areas and for detecting tuberculosis (TB) in humans. The organization began its research with Africa’s giant pouched rats (Cricetomys gambians) in 1998 and has since grown to over 200 employees across five countries. The organization's de-mining operations are currently active in Mozambique, Angola, and Thailand while TB screening operations are underway in Tanzania and Mozambique. APOPO’s Mine Action team has cleared a total of about 6 million square meters of land and APOPO’s TB detection rats have detected more than 3,000 cases of active TB missed by microscopists, who routinely screen for the disease in resource-poor countries. The success of APOPO’s rat detection technology has been widely recognized and in 2013 the organization was ranked by the Global Journal as the 13th best NGO in the world.
Although the rats do well, APOPO is constantly endeavoring to improve their performance and to develop new humanitarian uses for their scent-detecting abilities. In 2011, the author (hereafter "behavioral consultant") was hired by the organization to develop and implement procedures to improve the consistency and effectiveness of training procedures, with special emphasis placed on developing strategies for sustaining high-quality training without reliance on foreign, temporary staff. Historically, such individuals have regularly come and gone, to the detriment of the organization's effectiveness. In addition, the organization had reached a point in which several of the administrative staff felt the organization was ready to invest in new applications of the detection rats. Therefore, conducting publishable research investigating potential new applications was deemed highly important for determining which applications were most worthy of subsequent investments by the NGO as well as potential donor agencies.
EXPERIMENT ONE

Using Task Clarification and Feedback Training to Improve Staff Performance

Although the specifics differ, the fundamental approaches to improving the efficiency and effectiveness of any business are virtually the same. As Walsh and Lenihan (2006) argue, it is reasonable to assume that the same general strategies that improve the effectiveness of other organizations, such as for-profit businesses, would be of value in improving the effectiveness of NGOs. These practices of organizational governance are collectively referred to as "managerialism" in the international development literature. Despite the increasing pressure to adopt these practices and the general recognition of their benefits, managerialism is often met with some resistance from individuals working in the NGO sector.

Two commonly cited concerns involve the possible diversion from their value-driven goals as NGOs as well as the potential costs associated with the change in practices or staff required to manage those practices (Lindenberg, 2001; Meyer & Rowan, 1991; Roberts, Johnes & Frohling, 2005). According to Lindenberg (2001), "Those working in the NGO sector recognize that NGOs that do not adapt their strategies and promote greater impact, efficiency, and accountability run the risk of bankruptcy, as well as irrelevance. Yet they fear that too much attention to market dynamics and private and public sector techniques will destroy their value-based organizational culture" (p. 248).

Research investigating the use of managerialism in the non-profit sector is relatively new (within the last decade), and appropriate frameworks for adapting various organizational
strategies from business and industry are lacking (Mannell, 2010). OBM is one such strategy worthy of further investigation in the NGO setting.

OBM techniques provide a framework for identifying critical behaviors, determining environmental variables that affect these behaviors, and manipulating those variables to achieve specific business results crucial to an organization's stated mission. Several commonly used strategies characteristic of the field of OBM appear to be well suited for use in NGOs, in that they are effective, adaptable to various circumstances and relatively easy to implement at little cost.

One of the more popular strategies for enhancing employee performance involves the use of performance feedback. Although there is no consensus regarding how “feedback” should be defined, most definitions refer to feedback as involving information given to a person regarding their performance on a particular task in relation to a specific standard or goal (see Alvero, Bucklin & Austin, 2001; Prue, Frederiksen & Bacon, 1978). There is an extensive literature demonstrating the effectiveness of feedback interventions to improve performance in a variety of work settings including, but not limited to, the banking industry (Crowell, Anderson, Abel, & Sergio, 1988), hotels (LaFleur & Hyten, 1995), the food service industry (Austin, Weatherly, & Gravina, 2005; Pampino, Heering, Wilder, Barton, & Burson, 2003), manufacturing plants (Komaki, Barwick, & Scott, 1978) and mining (Fox, Hopkins, & Anger, 1987). Research has also shown that the manipulation of specific characteristics of feedback can alter the effects on performance. For example, immediate and specific feedback has been shown to be particularly effective (Balcazar, Hopkins, & Suarez, 1986).
Another powerful OBM intervention is task clarification. Task clarification was defined by Crowell, Anderson, Abel and Sergio (1988) as the "precise specification of behavioral components of a job" (p. 65), and can be used in the form of physical prompts such as checklists or job aids. Various forms of task clarification have been used successfully to improve work performance in many settings (see Anderson, Crowell, Hantula & Sioky, 1988; Anderson Crowell, Sponsel, Clarke & Brence, 1982; Bacon, Fulton, & Malott, 1982), however, it is most often used in combination with feedback or other types of intervention packages (e.g., Austin, Weatherly, & Gravina, 2004).

Developing and implementing job aids that provide task clarification and performance feedback is a prime example of the kind of OBM intervention that might benefit NGOs because research demonstrates that both can produce meaningful and sustainable improvement in workers' performance across a variety of settings (e.g., Austin, Weatherly, & Gravina, 2005; Balcazar, Hopkins, & Suarez, 1985; Gravina, VanWagner, & Austin, 2008; Prue & Fairbanks, 1981). There is, however, very little, if any, published research that could be reasonably construed as involving the use of OBM to benefit NGOs and to our knowledge none that demonstrates the value of task clarification or performance feedback for such applications. For example, a recent (November 4, 2012) search of the Journal of Organizational Behavior Management using the Scopus database and "nongovernmental organization," "non-governmental organization," "non-profit" and "NGO" as search terms failed to reveal a single relevant article. The purpose of this experiment was to investigate the feasibility of using a job aid (i.e., task clarification) in combination with supervisor training in delivering appropriate feedback in order to improve the performance of staff and supervisors.
A review of the mine detection rat training and accreditation process revealed a wide range in the number of days taken to internally accredit an animal for operational use, suggesting a high degree of variability in the performance of animal trainers. For example, in 2008 APOPO's trainers produced 38 rats adequate for operational use. On average, 252 training days were required for individual rats, but the range across rats was 164-590 days (Poling et al., 2010). Trainers consistently differ in how quickly they produce operational animals and it appeared that trainers using non-optimal training practices would benefit greatly from more effective and comprehensive supervision. If the variability in trainer performance can be reduced, then the overall process including time, money and other resources can be reduced as well, thereby improving overall efficiency.

In an attempt to reduce the variability in trainer performance, the performance of the supervisors was examined because research has shown that the skills of supervisory interactions with staff is an essential source of sustainable support for staff following training procedures (e.g., Clark, Wood, Kuehnel, Flanagan, Mosk, & Northrup, 1985). There was little evidence to suggest that the supervisors lacked the knowledge of animal training concepts or the ability necessary to deliver comprehensive supervision. Each supervisor at the NGO had, in the past, participated in an on-line course in animal training in addition to receiving a certification from an organization-sponsored training class designed to teach technical skills in conducting and evaluating behavioral training. However, informal observations of the supervisors' performance while evaluating the animal trainers suggested that their performance also varied dramatically from supervisor to supervisor. Given that there are a number of critical trainer behaviors required for
optimal training at any given stage, the most likely solution to assist with monitoring and enforcing these behaviors would involve precise specification (i.e., pinpointing) of each key behavior. Therefore, a task analysis was conducted and a job aid was developed to delineate the critical training behaviors required for optimal training at one specific stage of training. To supplement the use of the job aid, a brief training intervention was employed to improve the effectiveness of supervisor feedback on staff performance.
METHOD

Setting

Depending on the service that an individual rat will eventually perform (e.g., detecting landmines, detecting human tuberculosis), there are multiple training stages for the animal. For detecting landmines there are approximately nine stages that can be divided into two categories: the "pre-training" stages that take place indoors and the "advanced" stages that take place outdoors on a large training field. Prior to the study, a large number of rats were required to undergo remedial training as a result of skill deficiencies demonstrated at the advanced field training stages (e.g., walking in the wrong direction when crossing training boxes, responding to the wrong trainer after hearing the click). Given that these skills are specific skills trained at the stage immediately prior to the advanced field stages, a stage known as the "soil-floor stage", and the number of rats that had to be re-trained, the organizational staff suggested focusing the study on staff performance at this stage. The soil-floor stage is a pre-training stage, though training activities take place outdoors on a small 3 by 8 m training box, filled with local red soil, and serves as a transitory stage between the indoor and outdoor training stages. After observing staff performance at all stages using informal daily observation sheets, and noting sub-optimal aspects of trainer performance at the soil-floor stage, the behavioral consultant agreed with the supervisors regarding the importance of the soil-floor stage and selected it for the focus of this study.
**Participants**

Participants were recruited from the organization's workforce of approximately 50 in-country workers. Supervisors were eligible to participate if they held either a supervisory or managerial position as specified in their job description. Trainers were eligible to participate if they had been employed as an animal trainer at the organization for no less than one month and had rats about to begin training or were already training at the soil-floor stage. Trainers employed for less than one month would still be undergoing the organization's training procedures for animal trainers, not yet certified and were thus excluded from participation. All supervisors and trainers that met these criteria were eligible to participate in the study.

The Human Subjects Institutional Review Board of Western Michigan University approved the experimental procedures (Appendix A). Prior to participant selection, participants were told that the study was designed to evaluate methods for maintaining appropriate training behavior. Individuals who expressed interest in participating were given an informed consent document and, upon signing, were enrolled as participants to be randomly selected.

Three trainers (all men) and three supervisors (two men and one woman) participated in the study. All participants were between 18 and 40 years of age, and had 5-11 years of experience working for the organization. Five of the six participants were native to Tanzania, and the sixth participant was native to Kenya. Education backgrounds varied among the participants. Two of the participants had completed Form 4 of secondary school (typically for individuals 10 to 14 years of age) while another two had completed Form 6 (for individuals 14 to 16-years-of-age) and earned a Certificate of
Secondary Education. One participant held an online diploma in Animal Training, and one participant was enrolled in a Masters of Business Administration program at a local university at the time of the study. All six participants had earned internal organizational certificates in animal training provided by a leading animal training expert from Denmark. The certification process consisted of a 3-week class covering the conceptual basics in principles of behavior and training procedures outlined in the organization's standard operating procedures (SOP) manual. The workshop included both theoretical as well as hands-on practice and guidance from the instructor. The class was followed by a comprehensive exam. All participants demonstrated competency in English and fluency in KiSwahili. English was used throughout the study.

*Job Aid*

Initially, a task analysis was performed by the behavioral consultant to ascertain the trainer behaviors that were required to generate appropriate responding in the rats, which were being taught to sniff and scratch at small perforated aluminum containers ("tea eggs") that contained 2,4,6-trinitrotoluene (TNT), the primary explosive in most landmines. The task analysis was based on observations of ongoing training on the soil-floor coupled with thorough discussions with the supervisors and perusal of the organization's SOPs for training. Following the task analysis, a simple 11-item job aid was prepared (see Appendix B) to guide supervisors in monitoring the performance of trainers and prompt appropriate feedback when necessary. Nine items specified an action by the trainer, such as "the trainer stood outside the training box throughout the entire training session," or "the trainer practiced on both sides of the training box," and two of the items specified a *result* of trainer behaviors, specifically, "the rat appears
comfortable" and "the rat appears motivated." For simplicity, the items will be identified collectively as trainer behaviors.

Each item on the job aid included a brief description to help the supervisor evaluate the trainer's behavior, or the rat's performance, thereby, the trainer's behavior, correctly (see Table 1). For example, "the rat appears comfortable" included the following description: "The rat is sniffing and exploring the area freely; the rat is not stiff or jumpy." The supervisor's task was to respond to each item by indicating whether, for trainer activities, the listed task was performed, and for results of trainer activities, whether the rat was behaving as indicated. She or he did so by circling "Y" (yes) or "N" (no). Each item on the job aid was written in such a way that the "correct" answer, that is, the answer indicating appropriate trainer performance was "Y," and any "N" response indicated the occurrence of a training error. If a supervisor scored any item as "N," she or he was instructed to provide correct feedback, as described later, and to check the box labeled "FB" (feedback) to indicate that feedback was provided. In this way, the supervisor's performance was evaluated on the extent to which he or she verbally corrected the trainer following an observation of a training error - an essential behavior when engaging in appropriate supervision of staff.

Correct feedback statements were defined as "immediate, specific, and corrective." For feedback to be considered "immediate," the supervisor had to issue a verbal statement (feedback) within 10 seconds of the trainer error. To be considered "specific" and "corrective," the feedback statement had to state the error (e.g., "You clicked before the rat touched the target) as well as the appropriate behavior that the trainer should emit (e.g., "Next time, wait until you see the rat physically touch the target,
and then click as soon as possible”). The definition of "correct feedback statement" used in the present study was intended to ensure that feedback was both specific and timely; two characteristics of effective feedback (Balcazar, Hopkins & Suarez, 1986). Feedback in the form of praise was discussed and encouraged, but not required, to keep data collection simple. The consultant scored every item on the job aid each day of the study and the supervisors did likewise.

The job aid had three important functions. First, it clarified desired trainer (e.g., trainer stands outside of training box throughout session) and supervisor (e.g., specified when to provide feedback) behaviors. Second, it provided a means of quantifying these behaviors, and thereby the effectiveness of the intervention. Third, it provided a means of determining whether supervisors used the job aid appropriately. That is, by comparing the daily scores recorded by supervisors with the scores recorded by the consultant, treatment integrity could be calculated.

**Dependent Variables**

To quantify trainer performance, each training day a percent correct score was calculated by dividing the number of correctly completed training items by the total number of training items and multiplying by 100. To quantify the performance of supervisors, a percent correct score was calculated by combining the number of correctly completed training items plus the number of correct feedback statements delivered, divided by the total number of training items, and multiplying by 100. Treatment integrity was measured by calculating IOA between the scores recorded by the consultant and the scores recorded by the supervisor.
<table>
<thead>
<tr>
<th>Training Behaviors and Results</th>
<th>Brief Description</th>
</tr>
</thead>
</table>
| Type and position of Tea-Eggs is appropriate | If training #1: All or majority of tea-eggs are positive (see Appendix B)  
Buried at appropriate depth?  
Rat's 1st day = on top  
Rat's 2+ days of training = slightly covered  
Rat learning strong indications = buried deeper |
| The rat appears comfortable | The rat is sniffing and exploring the area freely; the rat is NOT scared or stiff |
| The rat appears motivated | After sniffing a positive tea-egg, the rat immediately scratches/digs/bites  
After hearing the click, the rat responds immediately by moving toward the trainer (or looking around as a reaction to the click)  
The # of responses to the click are greater than the # of no responses |
| Pulling the rope/rat | Short tugs are allowed when the rat has stopped to groom for greater than 15 seconds |
| Timing of the click is appropriate | When tea-eggs are on the surface - the click is timed with sniffing/contact with the tea-egg (not when rat is biting the tea-egg)  
When the tea-egg is buried, the rat scratches and the trainer clicks quickly |
| Trainer stepping into the training box | If training #3 - the trainer steps into the training box and gradually moves toward the side to train the rat to walk to the side of the box |
| The trainer practices on both sides of the box | If the rat has already learned to walk to one side of the training box for 5+ consecutive trials, the trainer moves to the opposite side to train the rat on both sides of the box |
| Trainer follows the "one direction" rule | Two trainers are training  
The rat walks in one direction across the lane.  
If the rat attempts to walk in the opposite direction, the trainer holds the rope tight (no pulling!). |
| The correct trainer clicks and rewards the rat | The trainer rewards the rat for responding to the person who clicked and does not give food if the rat returns to the wrong trainer (who did not click). |
| Measuring strength of indications | If the trainer is shaping strong indications, the trainer is counting out loud or using a stopwatch |
| The trainer is focused on the right skill for this rat | The rat demonstrates the skills listed above the skill being trained today (see Job Aid) |
**Procedures**

A multiple-baseline-across-subjects experimental design was used. In order to ensure consistency, each supervisor was randomly paired with one trainer throughout the study. Each supervisor-trainer dyad met at approximately the same time each day. Training sessions lasted 30 to 45 minutes each day and occurred 4-5 times a week (no more than once a day). Each trainer trained two or three rats each session. Throughout the study, the number of rats trained each day by a trainer and the specific animals trained did not change.

During all training sessions, data were collected by the behavioral consultant, who was a doctoral candidate in an Applied Behavior Analysis program, and occasionally by a research assistant, whose data were used in the calculation of inter-observer agreement (IOA) data. The research assistant was a Tanzanian rodent trainer with a masters' degree in Rural Development. Prior to the study, the consultant trained the research assistant to use the job aid during observations of training sessions until IOA between his score and the consultant's score exceeded 85% for three consecutive training sessions. During the study proper, the research assistant independently scored 17 of 51 training sessions (33%) across all phases and participants.

**Baseline**

During baseline, the trainers and the supervisors were asked to conduct their daily training or supervisory routines as usual. In other words, the supervisor was asked to "supervise" as usual, and the animal trainers were asked to train as they normally do. Only the consultant and the research assistant used the job aid to collect data on
supervisor and trainer performance, and neither interacted with the participants during the training sessions.

Training

Prior to the first animal training session, the consultant met privately with each supervisor and reviewed the job aid together. During this short informational session, each of the 11 items comprising the job aid were discussed, examples of using the job aid were provided, and the supervisors were instructed on how to deliver "correct feedback," as defined previously. Following this brief instruction, the supervisor was asked to score a completed job aid to demonstrate comprehension of the scoring procedures.

During the animal training sessions, the supervisor practiced using the job aid and delivering feedback. This way, the consultant and the supervisor scored both the trainer's and the supervisor's performance using the job aid. If the trainer emitted a training error, and the supervisor did not provide feedback, training was paused and the error and solution were discussed. If the supervisor attempted to deliver feedback but failed to include one or more of the three components of effective feedback, training was paused while the consultant reviewed the feedback statement with the supervisor and described the missing component(s).

Intervention

Following the 3-day training phase, the behavioral consultant did not interact with the supervisor during the training session and, although both the supervisor and the behavioral consultant scored the training session, their scores were not compared.
**Generalization and Maintenance**

Two types of follow-up data were collected following the intervention. First, approximately one week following the intervention's end, one to two sessions were observed to evaluate maintenance of the trainers' performance without the supervisors' presence. Second, as several of the training behaviors identified on the job aid also applied at the field training stage following the soil-floor stage, the trainers' performance in that field was briefly assessed (1-2 training sessions) using the job aid prior to and following the intervention. In this way, the extent to which the trainers changed their training performance across stages as a result of the intervention could be briefly assessed.

**Inter-observer Agreement**

Whereas treatment integrity was measured by calculating a percentage measure of IOA between the consultant's and the supervisor's scores on the job aid, a percentage measure of IOA between the consultant's and the research assistant's scores on the job aid was also calculated. The IOA percentage measure between the scores recorded by the consultant and the scores recorded by the research assistant was determined by calculating the number of sessions where their scores agreed perfectly (agreements), dividing that number by the total number of sessions (agreements plus disagreements), and multiplying by 100. Overall IOA was 95%. Given the high correspondence between the scores recorded by the consultant and the research assistant, only the former's scores are reported.

**Treatment Acceptability**

The acceptability of the job aid (i.e., its social validity) was evaluated using a modified version of the Treatment Acceptability Rating Form-Revised (TARF-R) originally developed by
Reimers and Wacker (1992). Following completion of the study, all three supervisors were asked to rate on a 1-5 scale (1 being the lowest and 5 being the highest score) the extent to which they understood and liked using the job aid as well as whether they would be likely to use the job aid during their supervisory routine in the future. The treatment acceptability rating form used in the present study is shown in Appendix C.
RESULTS

Treatment Effectiveness

Figure 1 shows the trainer and supervisor performance data (across rats) for each phase. During baseline, the average score for supervisors A, B, and C was 68.6%, 53.7%, and 49.7%, respectively. No trend in the data path was evident for any supervisor. The baseline phase lasted approximately 5, 8 and 9 sessions for Dyads A, B and C respectively. Performance improved during the training phase; the average scores in the training phase were 86.4%, 100% and 91.7% for Supervisors A, B, and C, respectively. Further improvement was evident for Supervisor A and C when the intervention phase was introduced, and decreased slightly for Supervisor B. During the intervention phase, the average performance scores for Supervisors A, B, and C were 98.4%, 98.5%, and 95.9%, respectively. The intervention phase lasted approximately 7, 9 and 5 sessions for Dyads A, B and C respectively.

Compared to baseline, trainer performance also improved during the training and intervention phase. During baseline, the average performance scores for Trainers A, B, and C were 66.8%, 45.5% and 49.0%, respectively. No trends in baseline performance were evident. During the training phase, respective average performance scores increased to 79.6%, 81.0%, and 86.3%. Finally, during the intervention phase, the average performance scores for Trainers A, B, and C further increased to 96.0%, 87.3% and 90.7%, respectively.

Table 2 shows the percent score of each supervisor on each of the 11 target behaviors collapsed across all sessions for each phase. The dashes in Table 2 indicate that the trainers engaged in the target behavior appropriately, leaving no opportunity for
Figure 1. Performance scores (percent correct responses) for trainers and supervisors (summed across rats) during each session of the study. Higher scores represent better performance.
Table 2

Percent scores for supervisors on individual target behaviors across each phase

<table>
<thead>
<tr>
<th>Phase</th>
<th>Supervisor A</th>
<th></th>
<th></th>
<th>Supervisor B</th>
<th></th>
<th></th>
<th>Supervisor C</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BL</td>
<td>TR</td>
<td>INT</td>
<td>BL</td>
<td>TR</td>
<td>INT</td>
<td>BL</td>
<td>TR</td>
<td>INT</td>
</tr>
<tr>
<td>1. Type and position of tea-eggs</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>50</td>
<td>-</td>
<td>100</td>
<td>17</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2. The rat appears comfortable</td>
<td>-</td>
<td>100</td>
<td>-</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>3. The rat appears motivated</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>42</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>4. Trainer never pulls rope/rat</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>33</td>
<td>100</td>
</tr>
<tr>
<td>5. Timing of the click is appropriate</td>
<td>8</td>
<td>78</td>
<td>100</td>
<td>12.5</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>50</td>
<td>-</td>
</tr>
<tr>
<td>6. Trainer never steps inside training box</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>7. Practice on both sides of training box</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>8. Trainer follows &quot;one direction&quot; rule</td>
<td>0</td>
<td>100</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Correct person to click</td>
<td>-</td>
<td>100</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Measuring strength of indications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Trainer is focused on training the right skill</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. Empty cells indicate a score could not be calculated because the target behavior was "not applicable", given the rat's level of training. Dashes indicate a score could not be calculated because there was no opportunity for corrective feedback.  BL = Baseline, TR = Training, INT = Intervention
the supervisor to give corrective feedback. Overall, scores on each target behavior increased from baseline to the intervention phase for all three supervisors. There was only one case in which the scores did not change from baseline (0%) to the intervention (0%) phase. This occurred for Supervisor C when evaluating the level of motivation exhibited by the rat. In the training phase, the motivation level was deemed appropriate and therefore no opportunity for feedback was recorded during this phase. In another case, there was no opportunity for Supervisor A to give feedback for the target behavior "type and position of tea-eggs" across all three phases because the trainer had demonstrated this behavior appropriately throughout each phase.

There were a few individual differences in performance across supervisors as far as which target behavior they were more likely to correct during baseline. Supervisor A provided the most corrective feedback during baseline on "timing of the click" (8%) and improved additionally in the intervention phase by correcting the trainer on 100% of the opportunities provided. Supervisor B and C provided the most corrective feedback during baseline when evaluating the type and positioning of tea-eggs used (50% and 17% respectively).

Table 3 shows the percent scores for each trainer on each of the 11 target behaviors collapsed across sessions for each phase. Overall, scores on each target behavior increased from baseline to the intervention phase for all three trainers. As shown in the Table, the first two target behaviors (type and position of tea-eggs, and comfort of the rat) and the last target behavior (trainer is training the correct skill for the rat) listed had the highest scores across all three trainers during baseline which suggests that these target behaviors were the least problematic for the trainers. The target behaviors that were demonstrated appropriately the least frequently during baseline involved "pulling the rope/rat" and "timing of the click". These two
target behaviors also remained the lowest scores during the intervention phase for Trainer A and Trainer B. Trainer C had trouble with "stepping inside the training box", demonstrating this target behavior appropriately for 6% of the baseline sessions but improved significantly by demonstrating the correct response for 90% of the intervention sessions.

Three of the target behaviors ("following the 'one direction' rule", "correct trainer to click and reward", and "measuring strength of indications") were indicative of the more "advanced" behaviors to train at this stage (see Appendix B for the order in which behaviors were specified to be trained). Only the rats belonging to Trainer A demonstrated the prerequisite behaviors necessary to be deemed ready for these three subsequent skills. Hence, these three target behaviors were evaluated only for Trainer A and not for Trainers B and C. The rats of Trainer B improved enough during the baseline, training and intervention phase that during the last three sessions of the intervention phase, the target behavior "practice on both sides of the training box" could be evaluated by the supervisor (i.e., the rat had learned to reliably return to one side of the box and was ready to learn to return to the other side of the box). Of those three training sessions, the trainer failed to practice on both sides of the training box the first time, was corrected by the supervisor on the second session, and practiced appropriately independently on the third session (33% score for the trainer and 50% for the supervisor). Trainer C did have rats that were ready for training on both sides of the training box and did so appropriately for 7%, 83% and 80% of the baseline, training and intervention phases respectively. None of the trainers' rats were deemed ready by the supervisors to be trained to indicate for a greater amount of time, and as such, none of the trainers were evaluated on this target behavior (i.e. "measuring strength of indication").
Table 3

Percent scores for trainers on individual target behaviors across each phase

<table>
<thead>
<tr>
<th>Phase</th>
<th>BL</th>
<th>TR</th>
<th>INT</th>
<th>MP</th>
<th>BL</th>
<th>TR</th>
<th>INT</th>
<th>MP</th>
<th>BL</th>
<th>TR</th>
<th>INT</th>
<th>MP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Type and position of tea-eggs</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>63</td>
<td>100</td>
<td>89</td>
<td>100</td>
<td>67</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2. The rat appears comfortable</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>63</td>
<td>67</td>
<td>89</td>
<td>100</td>
<td>94</td>
<td>100</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>3. The rat appears motivated</td>
<td>87</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>25</td>
<td>33</td>
<td>83</td>
<td>100</td>
<td>94</td>
<td>100</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>4. Trainer never pulls rope/rat</td>
<td>7</td>
<td>22</td>
<td>71</td>
<td>100</td>
<td>0</td>
<td>83</td>
<td>94</td>
<td>100</td>
<td>0</td>
<td>50</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>5. Timing of the click is appropriate</td>
<td>7</td>
<td>0</td>
<td>64</td>
<td>100</td>
<td>0</td>
<td>67</td>
<td>61</td>
<td>75</td>
<td>17</td>
<td>33</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>6. Trainer never steps inside training box</td>
<td>87</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>94</td>
<td>100</td>
<td>6</td>
<td>67</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>7. Practice on both sides of training box</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33</td>
<td>75</td>
<td>7</td>
<td>83</td>
<td>80</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Trainer follows &quot;one direction&quot; rule</td>
<td>17</td>
<td>67</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Correct person to click</td>
<td>100</td>
<td>78</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Measuring strength of indications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Trainer is focused on training the right skill</td>
<td>93</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>75</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Note. Empty cells indicate that a score could not be calculated because the target behavior was "not applicable", given the rat's level of training. BL = Baseline, TR = Training, INT = Intervention, and MP = Maintenance Probe.
**Generalization and Maintenance**

One week following the intervention, trainers sustained comparable performance. During the maintenance probe, Trainer A, B, and C demonstrated 100%, 92.9% (average across 2 days), and 100% correctly completed training items, respectively. As shown in Table 3, the target behaviors with the lowest scores during the maintenance probe were "timing of the click" (75%) and "practicing on both sides" (75%) for Trainer B. It is noteworthy that both of these scores were still improvements from the scores that Trainer B earned in the intervention phase (61% and 33% respectively).

Figure 2 shows scores for all three trainers at the second training stage prior to and following the intervention. For Trainers A, B, and C, the mean scores before the intervention were 68.8%, 70.8%, and 62.5%. Their respective scores after onset of the intervention increased to 93.8%, 100.0% and 93.8%.

![Figure 2](image.png)

**Figure 2.** Performance scores for all three trainers at a field training stage, where the intervention was not used, prior to and following the intervention.
Table 4 shows the scores for each trainer on the individual target behaviors at the advanced field stage prior to and following the intervention. Prior to the intervention, the target behaviors with the lowest scores involved "pulling the rope/rat", "following the one direction rule" and the "correct person to click". Following the intervention each of these scores increased. The majority of the other target behaviors had high scores prior to and following the intervention, which suggests that while different target behaviors may be problematic at different stages, improved supervision at one stage, could improve trainer performance across multiple stages.

Table 4

Percent scores for trainers on individual target behaviors at a secondary training stage prior to and post intervention

<table>
<thead>
<tr>
<th></th>
<th>Trainer A</th>
<th></th>
<th>Trainer B</th>
<th></th>
<th>Trainer C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>1. Type and position of tea-eggs</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>2. The rat appears comfortable</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>3. The rat appears motivated</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>4. Trainer never pulls rope/rat</td>
<td>0</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>5. Timing of the click is appropriate</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>6. Trainer never steps inside training box</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>7. Practice on both sides of training box</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>8. Trainer follows &quot;one direction&quot; rule</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
9. Correct person to click | 50 | 100 | 100 | 100 | 50 | 100
10. Measuring strength of indications | N/A | N/A | N/A | N/A | N/A | N/A
11. Trainer is focused on training the right skill | 100 | 100 | 100 | 100 | 100 | 100

Note. "N/A" means "Not Applicable".

_Treatment Integrity_

Figure 3 shows percent agreement (a measure of treatment integrity) between the scores recorded by the consultant and the scores recorded by the individual supervisors during the training and intervention phases. During the training phase, the average IOA for Supervisor A, B and C was 86.9%, 83.3% and 86.4%, respectively. Each supervisor demonstrated 95% or higher agreement by the third training session. The average IOA for Supervisors A, B and C during the intervention phase was 96.1%, 97.25 and 97.3%, respectively.

_Treatment Acceptability_

Figure 4 shows the responses of each supervisor to the five questions used to evaluate treatment acceptability. Overall, the results were positive. Using a 5-point Likert scale, with 1 indicating the least favorable rating and 5 the most favorable rating, the mean and mode for the five questions was 4. The 15 responses comprised four 3s, eight 4s, and three 5s. No score lower than "neutral" (i.e., 3) was obtained. Item 1, concerning the clarity of understanding for using the job aid, had the highest ratings (scores 5, 4 and 5, whereas items 2, 3 and 5 had the lowest ratings (scores 3, 4 and 4).
Figure 3. Percentage of agreement (a measure of treatment integrity) between the scores recorded by the consultant and the scores recorded by the individual supervisors during the training and intervention phases. Higher scores indicate better treatment integrity.
Figure 4. Responses of each supervisor to the five questions used to evaluate treatment acceptability. Higher ratings indicate more favorable responses whereas lower ratings indicate less favorable responses.
DISCUSSION

The purpose of the experiment was to develop and utilize a relatively simple intervention package for producing and maintaining desirable staff and supervisor behaviors. A multi-function job aid was devised to meet this objective in conjunction with a brief training session on how to deliver effective feedback. The results suggested that the intervention improved both the performance of the supervisors as well as the trainers. Performance of all three trainers and all three supervisors was sub-optimal during baseline but improved substantially when the training phase was introduced, and further improved during the intervention phase, in which both trainers and supervisors performed at a consistently high level. Maintenance and generalization data, although limited, suggest that the intervention produced enduring beneficial changes in the behavior of trainers, and that these effects were evident without the supervisor's presence as well as beyond the experimental context. These results suggest that incorporating the job aid into the training system will help maintain optimal training practices without requiring daily supervision. Moreover, social validity data suggest that supervisors viewed the intervention as highly acceptable and easy to use. These data suggest that trainers and supervisors are likely to continue use of the job aid and the treatment integrity data suggest that they will be able to do so effectively.

As noted, the job aid was intended to serve multiple functions including data collection, task clarification, and the provision of appropriate feedback to trainers from supervisors. Because the supervisors completed the form every training day and had to record their own behavior (i.e., whether feedback was delivered for training errors), the
job aid also gave them feedback regarding their own performance, with higher scores indicating better performance. Although simple, the job aid is in a real sense a package intervention. That being the case, it is impossible to specify the mechanism or mechanisms responsible for the results. It is, of course, likely that both task clarification and feedback played a role, because both strategies, alone and in combination, are often effective in producing desired changes in behavior (e.g., Austin et al., 2005; Balcazar et al., 1985; Gravina et al., 2008; Prue & Fairbanks, 1981). It is noteworthy that the supervisors could have recorded scores more favorably for themselves in order to obtain higher performance scores, however, the high agreement shown by the treatment integrity data suggests that they did not do so. In addition, the supervisors in the present study quickly learned how to use and score the job aid, requiring no more than 2-3 training sessions, and consistently delivered immediate corrective feedback after such training. However, although performance improved on the first training session at the start of the intervention, it is unclear whether use of the job aid without a similar feedback training procedure with the consultant would yield comparable results. A follow-up investigation examining the extent to which the current supervisors can train additional supervisors on how to use the job aid and deliver feedback successfully would affirm the tool's long-term value.

Although the cause of the staff performance deficits were not directly assessed, a significant weakness in the present study, there was evidence to suggest that the low performance scores during baseline were not a result of knowledge or skill deficits. First, prior to the start of data collection, a staff meeting was held to review the organization's standard operation procedures (SOPs) for training at the soil-floor stage and included a
discussion of the proper steps to take for training specific animal behaviors. The review was approximately an hour and a half long and included questions and answers between the animal trainers, the supervisors and the consultant, though mostly led by the consultant and one supervisor. In addition, trainers identified common problems experienced at the soil floor and provided tips to one another for addressing specific issues. For example, training the rat to return to one side of the training box was discussed and identified as an important pre-requisite for training the rat on both sides of the training box. Secondly, both the trainers and the supervisors had all worked for the organization for 5-11 years, had been certified internally as animal trainers, and were more than willing and able to describe the rules involved in animal training. In addition, at a recent staff meeting (approximately one month before the start of the study)- the problem of "pulling the rope/rat" had been discussed among the trainers and upper management as well as methods for minimizing the problem - and yet, this was still one of the most frequently occurring trainer problems observed in this study, and often neglected by supervisors. We believe that the most beneficial aspect of the job aid was its comprehensive nature and interactive design. The job aid was comprehensive in that it listed all of the critical behaviors necessary for engendering appropriate responding in the rats, and included descriptions to help the user distinguish instances of correct or incorrect trainer behaviors. It was interactive in that the supervisor was asked to actively distinguish instances of correctly performed target behaviors and mark whether feedback was delivered when appropriate to do so. In fact, one of the reasons that the target performance for the supervisors was so narrowly defined (i.e., restricted to simply "provide immediate, specific and corrective feedback" rather than "be on time, use
checklist and pen, correct and praise trainer") was the observation that supervisors were very quiet when "supervising", and secondly, inconsistent within and among themselves. This observation was supported by the supervisors' performance scores during baseline in which corrective feedback was provided inconsistently across target behaviors as well as infrequently. It is possible that the task clarification function of the job aid was not necessarily as much as a benefit for the trainers as it was for the supervisors in that it required them to actively distinguish correct and incorrect trainer behaviors and deliver correct feedback. Their performance score was a clear result of the extent to which they performed this task appropriately, thereby potentially creating sufficient motivation to do so consistently. However, the lack of an explicit assessment of the initial performance deficiencies was a clear weakness and should be included in subsequent research.

Similarly, research investigating the utility of a job aids independent of feedback training for supervisors would help clarify the circumstances in which each might be more useful.

One concern with research, especially in organizational settings, is the potential reactivity of subjects to observers (Reid, Parsons & Green, 1989). In the present study it was possible during any given training session for as many as three people to be observing the training session and taking notes. However, the possible reactivity to multiple observers was not considered to be a major concern in this experiment because the presence of observers taking notes and discussing trainer or rat performance was not an unusual or new occurrence in any way for the trainers or the supervisors. First, both the trainers and the supervisors were accustomed to the consultant's presence in observing sessions daily, taking notes, asking questions or making suggestions regarding rat or human performance. This was true for all stages of training including the soil floor
and the advanced field training stage where the generalization data were collected. The consultant had been analyzing and collecting data on different aspects of human or system performance at the organization for over a year at the start of this study. The trainers were also used to supervisors, even several at one time, observing their performance daily and providing feedback on their performance or the performance of the rat. Secondly, the subjects were accustomed to participating in experiments regularly, albeit in different ways, because much of their daily work involved empirically investigating and developing new procedures in order to improve animal performance or organizational procedures. Finally, the low performance scores during baseline suggest that both the trainers and the supervisors did not alter their behavior only in the consultant's presence.

Although the job aid was simple and effective, which are points in its favor, the most interesting aspect of the present study is not the intervention, but rather the setting in which it occurred, that being an NGO working in one of the poorest countries on Earth, one where the per capita gross domestic product from 2007-2011 was roughly $500. For comparison, the corresponding value for the United States of America was just under $50,000 (The World Bank, 2012). Conducting research intended to evaluate OBM interventions that are designed to improve organizations located in such settings poses significant challenges and, depending on how broadly one construes "OBM interventions", few, if any, relevant studies have appeared. One case study described the implementation of a business tool, the ISO 9000 Quality Standard in an NGO, the Cambodia Trust, and although the authors reported an "increase in effectiveness" and general acceptability by the staff, no outcome data were reported (Walsh & Lenihan,
In a more recent study, Muianga and colleagues (2012) combined the use of a staff checklist, formal training workshop, and the development of a manager evaluation survey to promote changes in work practices that would reduce workers' exposure to silica dust in a small-scale demolition site located in Maputo City, Mozambique. Results from a survey showed significant improvement in work practices following the intervention compared to baseline levels, but experimental control was not documented. The present study builds on this seminal work by showing that it is possible to arrange conditions so that relatively good experimental control is possible.

It must be emphasized, however, that conducting systematic research in a multi-cultural, third-world environment was far from easy with respect to the present study and what appears to be a short and simple investigation posed substantial practical and logistical challenges. For example, limited maintenance and generalization data on the supervisors' performance were collected because organizational needs and resources changed unexpectedly over a span of two weeks. First, as the experiment proper was coming to the end, a large number of rats were suddenly needed for operational demining, which required the supervisors to shift their attention to end-stage training of field rats and final testing procedures of those rats prior to their shipment to operational demining sites. Secondly, following the accreditation tests (the organization's formal testing procedures for determining a rat's capability to work in real land-mine fields), two of the supervisors took their annual one-month holiday leave without more than a week's notice. Another weakness is that the treatment acceptability data are hard to interpret. In part, this is due to the fundamental nature of such data, which are not direct measures of any important aspect of current or future behavior and are subject to influence by a wide
range of variables (Carter, 2008; Finn & Sladeczek, 2001; Lennox & Miltenberger, 1990). One such variable is likely to be the culture of the respondent, which can influence how employees interpret and respond to work-related questions (Riordan & Vandenberg, 1994). From our own experience and from general descriptions of the culture (e.g., Otiso, 2013), expressing strong feelings of any sort is not highly regarded among Tanzanians, nor is bearing bad news or disappointing a person of high status. Insofar as the behavioral consultant was an expatriate member of the organization's management team, the supervisors probably viewed her as a high-status person and for that reason may have responded favorably to the questionnaire, irrespective of their real feelings. On the other hand, a general reluctance to provide strong answers (i.e., 5s on our questionnaire) may have contributed to a paucity of such scores, even if the intervention was highly favored. Regardless, the extent to which the supervisors continue using the job aid and engendering high-quality performance, not their responses to any questionnaire, will be the ultimate measure of its worth.
EXPERIMENT TWO

Investigating the Olfactory Detection of Salmonella Bacteria in Horse Fecal Samples

Nosocomial infections (infections acquired in a hospital) are a major cause of morbidity and mortality in human and veterinary patients worldwide. In humans, diseases acquired in United States hospitals result in over 90,000 deaths per year and have an annual cost estimated to be in excess of $4.5 billion (Weinstein, 1998). Rapid detection of life-threatening nosocomial pathogens, such as Clostridium difficile and methicillin-resistant Staphylococcus aureus (MRSA) in human hospitals, and Salmonella in veterinary hospitals, has the potential of allowing better segregation of patients and improving the ability of hospitals to make immediate biosecurity decisions that could significantly reduce the risk to both patients and personnel.

Salmonella increasingly compromises human and animal health around the world (Mead et al., 1999; Voetsch et al., 2004). The World Health Organization (2005) estimates 1.4 million non-typhoidal Salmonella infections occur annually in humans, resulting in 15,000 hospitalizations and 580 deaths in the United States alone. In the past decade, salmonella has become the third-most-common cause of death of HIV-positive people in sub-Saharan Africa and is now recognized as an emerging and dangerous tropical disease (Feasey, Dougan, Kingsley, Heyderman, & Gordon, 2012). Salmonella is also a problem in veterinarian medicine. For example, a recent nosocomial salmonellosis outbreak at a large veterinary teaching hospital resulted in a high case fatality rate for patients, as well as a 10-month hospital disruption and total costs of $4.12


million for decontamination, remediation, and lost revenue (Dallap-Schaer, Aceto, & Rankin, 2010).

Although polymerase chain reaction (PCR) analysis of material from fecal culture remains the gold standard for the detection of *Salmonella* (Jones, 2008; Ward, Alinovi, Couëtil, & Wu, 2005), this method has a reported sensitivity of only 30-60% in clinical samples (Aceto et al., 2008) and is time consuming and expensive (Aceto et al., 2008; Ward et al., 2005). Therefore, a fast and inexpensive alternative to this technique for detecting *Salmonella* would be highly useful. Given their ability to detect the scent of *M. tuberculosis* in human sputum (e.g., Mahoney et al., 2011, 2012; Poling et al., 2010a), it is possible that pouched rats can also detect the scent of *Salmonella*. Experiment 2 will examine whether pouched rats can detect the presence of *Salmonella* inoculated in horse fecal samples.
METHOD

Subjects

Six giant African pouched rats (*Cricetomys gambianus*), four females (Femi, Fenella, Ghita and Rotarat) and two males (Geoff and Costanze), each approximately 2 years old, served as subjects. All six were from APOPO’s breeding colony, weaned at 4 weeks of age, and socialized by the APOPO staff during weeks 4-8 as described elsewhere (Poling et al., 2011). Femi and Fenella were housed together as were Geoff and Ghita. Costanze and Rotarat were housed alone. Food (other than pellets used during training) was restricted to daily feeding of two to three commercial rat pellets (Noyes Precision 90 mg Food Pellets with banana flavoring) one hour after each daily session and a Friday post-training meal consisting of half a tomato, half a banana, one slice of carrot, one quarter of an apple, and a quarter cup of peanuts. All rats had unlimited access to water. Each rat was weighed every Monday, Wednesday, and Friday to ensure that significant weight loss did not occur. APOPO has Animal Welfare Assurance to conduct this research (no. A5720-01) from the Office of Laboratory Animal Welfare, National Institute of Health, USA.

Previous Training History

The rats had a previous training history that was compatible with the present experiment. Prior to the present study, operant conditioning procedures described in detail elsewhere (Poling et al., 2010, b) were used to train each rat to detect 2,4,6-trinitrotoluene (TNT), the primary explosive in most landmines. In essence, the training involved socializing the rats through intensive contact with humans while they were 4-6 weeks old, establishing a click sound as a conditioned reinforcer, and providing
differential reinforcement for pausing for 3 s above holes below which samples containing TNT were present, but not for pausing above holes where the samples did not contain TNT.

**Apparatuses**

Experiment 2a occurred in a square metal cage (66 x 66 x 35 cm) with three sniffing holes (2 cm in diameter and 10 cm apart) centered in the floor (see Appendix D). A small plastic pot containing an appropriate training sample was placed immediately below each hole. At least one pot per trial, with its location selected at random, contained *Salmonella*-positive culture broth and the other two contained *Salmonella*-negative culture broth. If the rat emitted an indication response (i.e., paused for at least 3 s) over a positive sample, the trainer pressed a small handheld device to make a “click” sound, and presented a plastic syringe filled with 20 cc of banana mixture with crushed rat chow through a hole in the wall of the test cage.

Experiments 2b and 2c took place in a semi-automated cage (see Appendix D). The semi-automated cage was designed by APOPO in an attempt to minimize human errors and make the trainer’s task easier. Figure 5 shows a picture of the semi-automated cage in the top panel and a picture of a rat indicating on a sample in the lower panel. A second (1.08 meter in diameter) stage below the main stage contains a 3.5 cm wide rack used to hold the sample pots in the order in which each will be presented to the rats. The bottom stage turns in carousel fashion (manually) in either direction allowing the trainers to easily reach each sample pot and to avoid accidentally selecting the same sample twice. The upper stage consists of an aluminum floor, approximately 1 meter in diameter, and is encased by a plexi-glass wall. The walkway itself is 15 cm wide and 20
cm tall, with a perforated metal lid that can be lifted open on either side. Inside the walkway are two sniffing holes directly opposite the other and two parallel pellet dispensers perpendicular to the sniffing holes (as shown). The rats walk clockwise around the circular walkway pausing to sniff each hole. After the rat sniffs the first hole, the hole is then closed off by sliding, metal doors on either side of the sniffing hole. The trainer closes the sliding doors by pulling the metal sliding bar that is directly below the sniffing hole; this allows him or her to change the samples. In other words, as the sliding bar is pulled out to change samples, the sliding metal doors are simultaneously pulled closed. Likewise, once the samples have been changed, the sliding bar can be pushed back below the sniffing hole and the doors slide open simultaneously.

**Figure 5.** The semi-automated cage is shown in the top panel and a rat indicating on a sample in the lower panel. Notice the dual stages—the top for the rat to pass through, and the lower to hold samples.
Next to both sliding bars is a push button to activate the pellet dispenser as well as a green indicator light. Once a positive sample is set in the sniffing position, the trainer pushes the button to activate the pellet dispenser which releases 3 pellets if the rat’s nose interrupts the beam of light inside the rim of the sniffing hole for 3 seconds or more. Any time the rat’s nose interrupts the beam of light inside the sniffing hole for more than 3 seconds, the green light also illuminates to signify the indication. Thus, false indications can be identified. In addition, any time the rat makes an indication on either side of the carousel, the green light flickers on the opposite side to signal the indication to the other trainer. This way, the trainers and the note-taker know when an indication is made on either side of the apparatus without having to move.

Sample Preparation

Freeze dried *Salmonella enterica* serovar Newport (ATCC 6962), Gallinarum (ATCC 4961), Agona (ATCC 51957) and Saintpaul (ATCC 9712) were obtained from ATCC, USA and grown in nutrient broth to a bacterial load of approximately $10^9$ colony forming units per ml. Fifty vials containing sterilized nutrient broth (approximately 5mL) were then prepared; 20 vials were inoculated with the target bacteria and 30 control vials were not inoculated. After overnight incubation, bacterial growth appeared on the inoculated target material, but no growth was observed on the control substance. To ensure the growth of actual and only *Salmonella* bacteria, the target material was then inoculated onto *Salmonella-Shigella* agar and biochemical tests were completed using sucrose, lactose, mannitol, citrate, urea and triple sugar iron agar. All tests confirmed the growth of *Salmonella* bacteria. The target broth and pure control broth were then heat inactivated at 90°C for 2 hrs and stored at 4°C. Save for not having *Salmonella* added, the
control broth was prepared in an identical manner as to the target broth with respect to preparation, sterilization, incubation, inactivation, and storage.

Samples for Experiment 2a were prepared by placing five drops (150-200 µL) of either the inactivated bacteria (target) or pure broth (control) into a plastic sample pot. In the second and third phase, dried and fresh horse manure were added, respectively. A bucketful of freshly laid horse manure was collected from the horse pen of the Faculty of Veterinary Medicine at Sokoine University of Agriculture (Morogoro, TZ) and sun-dried for four days. Seven horses, two male and five female, ranging from 5 to 19 years of age, occupied the pen; the horse(s) producing the dung was not identified.

Prior to use, the horse manure was analyzed for the presence of *Salmonella* bacteria by adding 100 mg dried manure to 5 µL of selenite broth. The broth was incubated at 37°C for 8 hrs, and then inoculated onto *Salmonella*-Shigella agar. There was no growth observed after 24 hrs, confirming the absence of any *Salmonella* species. To prepare *Salmonella*-positive samples, five drops of *Salmonella* culture broth were added to 0.2 g dried horse manure and *Salmonella*-negative samples were prepared by adding the same amount of control broth to 0.2 g dried horse manure. For the third phase of Experiment 2a, the same amount of target or control broth was combined with 1 g of fresh *Salmonella*-free horse manure collected each morning immediately prior to behavioral testing. The manure was crushed, homogenized, and immediately used to prepare training samples. A small portion of the manure was set aside for bacteriological analysis in which approximately 1 gram of fresh manure was placed in sterilized selenite broth for enrichment and after 8 hours of incubation at 37 °C, it was inoculated into Xylose Lysine Deoxycholate agar which was maintained at 37° C for 24 hours. No
bacterial growth characteristic of *Salmonella* was observed on samples collected from two stalls, but there were ambiguous indications of growth for samples collected from two other stalls. In the latter cases, biochemical tests with urea followed by triple sugar confirmed the absence of *Salmonella* in the samples. For Experiment 2b and 2c, dried horse manure was used as described previously.

*General Procedures*

All sessions occurred once daily, five days a week (Mon-Fri) at approximately the same time each day (10:30 to 11:30 AM). Between all sessions, the apparatuses were wiped clean with 70% ethanol.

*Data Analysis*

For any diagnostic technique, both sensitivity - the ability to detect the presence of the pathogen (or disease) of interest - and specificity - the ability to detect the absence of the pathogen of interest - are important. Percent hits and percent correct rejections are conventional measures of sensitivity and specificity, respectively. Sensitivity and specificity were calculated for each session throughout all of the experiments, visually analyzed, and in some cases analyzed by statistical analyses using Minitab 15 Statistical Software.
Method

Training occurred manually in a square chamber and consisted of three phases. First, samples were prepared with five drops (150-200 µL) of either the target or culture broth alone. The acquisition criterion for the first phase was at least 60% sensitivity and specificity (greater than chance levels) for two consecutive sessions with the exception of the first session of discrimination training. In the second and third phase, samples were prepared using dried and fresh horse manure, respectively. During these two phases, the learning criterion was at least 85% group sensitivity and specificity for three consecutive sessions. Sessions comprised 30 trials with at least one positive and two control samples presented each trial.

Training involved establishing the scent of *S. enterica ser. Newport* as a discriminative stimulus for correct indications (pausing) by reinforcing with food pauses of at least 3-s duration that occurred at samples containing the bacteria, but not at samples that did not contain it. Indication responses above *Salmonella*-positive samples were followed immediately by a "click" and delivery of food, whereas responses above *Salmonella*-negative samples had no programmed consequences. Correct indications were recorded as “hits” whereas indications over negative samples, which had no programmed consequences, were recorded as "false alarms." If a rat sniffed a positive sample and did not emit an indication response, the trainer recorded the response as a “miss” and moved to the next trial. If the rat sniffed a negative sample and did not emit an indication response, the trainer recorded a "correct rejection." Hits and correct rejections are correct responses whereas false alarms and misses are errors.
Results

Figure 6 shows the average group performance across the three phases. All rats met the learning criterion for the initial odor discrimination within four sessions of the first phase when the target or control culture broth was presented alone. The group mean sensitivity and specificity during the first phase was 68.0 (range: 40 to 83%) and 74.3% (range: 42 to 100%), respectively. In the second and third phase, the learning criterion was met after 10 and 5 sessions respectively. For the last three sessions in the second phase the group mean sensitivity and specificity was 91.1%, 94.4%, 93.0% and 95.0%, 95.5%, and 96.3%, respectively. For the last three consecutive sessions in the third phase, the group mean sensitivity and specificity was 94.4%, 91.7%, 94.4% and 97.2%, 97.9%, and 97.2%, respectively.

Figure 6. Average sensitivity and specificity data for the rats as a group in the culture broth only phase (A), the dried horse manure phase (B) and the fresh horse manure phase (C).
Experiment 2b

Method

Once subjects met the learning criterion in Experiment 2a, discrimination training continued as described previously, except training occurred in the semi-automated cage. Sessions consisted of the presentation of 50 samples, 9 *Salmonella*-positive (18%) and 41 *Salmonella*-negative, in random order.

An A-B-A design was used to assess the rats’ performance in which phase A consisted of presentations of *S. enterica ser.* Newport and phase B consisted of presentations of *S. enterica ser.* Gallinarum. Phases were changed once the rats demonstrated stable performance using a moving mean criteria of sensitivity and specificity across three consecutive sessions with no more than 5% change or a consistent increasing or decreasing trend in means.

In order to approximate conditions in Experiment 2c, in which test samples would not be reinforced, an average of 7.5 of 9 (83%) (range 6-9) responses to *Salmonella*-positive samples were reinforced during the training procedures in Experiment 2b. Therefore, there was no change in the overall amount of reinforcement a rat received between Experiments 2b and 2c.

Double-Blind Tests

In order to minimize incidental trainer-rat cues, the trainers were blind to the true status of all samples during training in the semi-automated cage. When the rat indicated for 3 s or longer, the note-taker verbally instructed the trainer to reinforce by pressing a button to deliver the click and food reinforcer, or ignore the rat's response. In addition, three brief, double-blind tests were conducted to ensure the absence of cueing at different
points throughout the study. The first double-blind test occurred in the middle of the first phase in the semi-automated cage with the presentation of *S. enterica ser.* Newport samples and the second occurred during the second phase with the presentation of *S. enterica ser.* Gallinarum samples. The final double-blind test occurred at the end of Experiment 2c.

During the double blind tests, 1-2 samples marked as positive on the note-taker's data sheet contained control broth only, and therefore were false positives (i.e., true negatives). Simultaneously, 1-2 samples marked as negative on the note-taker's data sheet actually contained the target broth and, thus, were false negatives (i.e., true positives). If the rats were responding to human cues, responding would occur on the false positives and not on the true positives. On the other hand, if the rats were responding only to the target stimulus, responding would occur on the true positives and not on the false positives.

**Results**

Figure 7 shows the rats' performance as a group with *Salmonella* Newport or Gallinarum samples presented. The first phase took approximately 14 sessions, whereas the second and third phase took 21 and 9 sessions, respectively. In the first phase, when samples containing Newport were presented, the group average sensitivity and specificity was 74% (range: 59% to 89%) and 99% (range: 95% to 100%) respectively. The average group sensitivity increased to 88% (range: 72 to 98%) when the second strain was presented, and average specificity decreased to 93% (range: 87 to 100%). In the third phase, when the Newport samples were re-introduced, the group's average sensitivity remained at approximately 87% (range: 78 to 100%) and specificity increased to 97% (range: 95 to 99%), respectively. Performance remained stable during the double-blind
tests in which sensitivity and specificity was 80% and 100%, respectively, on the first test and 94% and 89% on the second test.

A one-way repeated measures ANOVA was conducted to evaluate sensitivity across the three presentations of the two *Salmonella* strains. Three levels of the independent variable were evaluated: *Salmonella* Newport initial presentation, *Salmonella* Gallinarum, and *Salmonella* Newport second presentation. Results indicated a significant effect, with $F(2, 10) = 14.41, p < .05, \eta^2 = .0785$.

Three pairwise comparisons were conducted to compare the means in sensitivity during each phase of *Salmonella* presentation. Two of the pairwise comparisons were significant, controlling for familywise error rate across the three tests at the .05 level, using the Holm's sequential Bonferroni procedure. The smallest $p$ value was for the comparison in sensitivity during the first presentation of *Salmonella* Newport samples.

**Figure 7.** Average sensitivity and specificity data for the rats as a group with *Salmonella* Newport (A) or Gallinarum (B) samples were presented. Data from the double-blind tests are shown in between the dashed lines.
and the second presentation of such samples, with a $p$ value of .002, which is less than $\alpha = .05 / 3 = .017$; therefore, the difference between the means is significant. Similarly, there was a significant difference between the means during the first presentation of *Salmonella* Newport and *Salmonella* Gallinarum, with a $p$ value of .023, which is smaller than $\alpha = .05 / 2 = .025$. The $p$ value of .28 for the comparison of means during the presentation of *Salmonella* Gallinarum samples and the second presentation of *Salmonella* Newport samples was not smaller than .05, and therefore not significant.

The results of a one way repeated measures ANOVA evaluating specificity levels across presentations of the two strains also indicated a significant effect, with $F(2, 10) = 47.57, p < .05, \eta^2 = .908$. Three pairwise comparisons were conducted and two were statistically significant controlling for familywise error rate across the three tests at the .05 level, using the Holm's sequential Bonferroni procedure. The smallest $p$ value was for the comparison of means in during the first presentation of *Salmonella* Newport samples and *Salmonella* Gallinarum samples, with a $p$ value of 0.00, which is less than $\alpha = .05 / 3 = .016$, and thus statistically significant. The second statistically significant difference in means was between Gallinarum samples and the second presentation of *Salmonella* Newport samples, with $p = .001$, which is less than $\alpha = .05 / 2 = .025$. The comparison of means for the first and second presentation of *Salmonella* Newport samples was not statistically significant ($p = .456$).
Experiment 2c

Method

Training continued in the semi-automatic cage as previously described, however, 4-5 positive samples containing either *S. enterica ser.* Newport or *S. enterica ser.* Gallinarum were presented (total 9 positive samples). Four test sessions, randomly scheduled, were conducted in which 2 of the 9 positive samples presented contained either *S. enterica ser.* Agona or Saintpaul. During these test sessions, samples containing either Newport or Gallinarum were thus reduced to 3 or 4. In all, 4 samples were presented across 2 test sessions for each of the novel targets. Indications on the samples containing the novel targets were not reinforced.

Results

Figure 8 shows the group's average sensitivity for Newport and Agona samples as well as the presentation of two novel *Salmonella* strains -Agona and Saintpaul. Overall, the group average sensitivity on samples containing *Newport* was 88% (excluding the last double-blind test data) and for Gallinarum 97%. Specificity was 97%. The overall group average sensitivity for samples containing Agona was 96% and for Saintpaul 71%. During this experiment, sensitivity and specificity during the double-blind test was 98% for both measures.
Figure 8. Average sensitivity data for the rats as a group when presented with Newport, Gallinarum (GLT) simultaneously as well as two novel strains - Saintpaul and Agona. Data from the double-blind tests are shown in between the dashed lines.
DISCUSSION

The present data show that with little training pouched rats accurately detected horse fecal samples inoculated with *S. enterica ser.* Newport. In addition, following discrimination training with a second strain, Gallinarum, both sensitivity and specificity improved when presented with the initial strain of *Salmonella* (i.e., Newport). When random, non-reinforced samples of two novel strains were presented, sensitivity and specificity remained high, although dropped slightly for *Salmonella* Saintpaul. These data show that rats can discriminate between *Salmonella* and other organism-infused equine fecal flora samples, and will do so with generally high sensitivity and specificity with some untrained strains of the bacteria. These results suggest that it is reasonable to pursue the possibility of using pouched rats to screen for the presence of *Salmonella* in veterinary hospitals and elsewhere.

One major advantage of using scent-detection rats for the identification of *Salmonella* is their ability to evaluate numerous samples in a short amount of time. In the present study, the evaluation of 50 samples took approximately 12.5 minutes, with a range of 8 to 17 minutes. Research with TB-detecting rats indicates that an individual rat can easily evaluate hundreds of sputum samples per day (Weetjens et al., 2009) and fecal samples could be evaluated with similar speed. As previous research has demonstrated, frequent sample collection and evaluation is necessary in order to minimize the chance that a sample is falsely evaluated as *Salmonella*-negative due to various reasons such as the dilution of the organism or problems with the transportation or storage of the samples (Amavisit et al., 2001). By using a highly sensitive and fast detection technology, which
scent-detection rats may prove to be, evaluating multiple samples from the same patient requires little extra work and expense.

To the authors’ knowledge, this is the first report of using any animal to detect the presence of *Salmonella* in horse fecal samples and its findings suggest that pouchcd rats can readily do so. Of course, the results of the present study are preliminary and further research, including studies of the rats’ ability to detect various strains of salmonella in samples from infected animals, humans and food products, is needed to ascertain whether scent detection rats provide a realistic option for *Salmonella* detection in various settings.
EXPERIMENT THREE

Training Rats to Find People

Human acts of war and terrorism, natural disasters, engineering mistakes, and inadequate maintenance can cause buildings and other structures to collapse, potentially trapping people under debris. Rapidly locating, extricating, and treating these disaster victims greatly increases the likelihood of their survival. Currently, organized search and rescue techniques rely primarily on human search teams that use visual and auditory cues to find victims (Wong & Robinson, 2004). Specially trained dogs, which search large areas for human scent and emit an indicator response, such as sitting or barking, when human scent is located, are also widely used for this purpose (Chiu et al., 2002; Snovak, 2004; Wong & Robinson, 2004). Although they are useful, scent detection dogs can take years to train, are inconvenient to transport, and are limited in their capacity to localize the source of human scent rising from rubble (Chiu et al., 2002; Wong & Robinson, 2004). Moreover, because of their size, dogs typically cannot penetrate rubble, but are limited to searching for scent as they move over and around debris (Federal Emergency Management Agency, 2010).

As emphasized at a workshop hosted by the United States Department of Justice, no proven technique is available for searching through and under debris without the time-consuming process of manual or mechanical rubble removal (Wong & Robinson, 2004). One recommendation made by workshop attendees, which included representatives of urban search and rescue teams and other first responders, was to investigate the possibility of training search animals smaller than dogs, specifically rodents and insects,
to find living disaster victims by moving through and under debris. A pilot project was recently initiated to explore whether giant African pouched rats (*Cricetomys gambianus*) could be trained to search for people and to return to their release point upon hearing a return signal. The present study reports these initial findings. Pouched rats, used successfully to detect landmines and human tuberculosis (e.g., Mahoney et al., 2011, 2012; Poling et al., 2010a, 2010b, 2011), are long-lived rodents native to sub-Saharan Africa. They are nocturnal, burrowing animals that are large enough to carry a small amount of weight, such as a backpack equipped with a camera and other detection or communication apparatuses. Therefore, assuming that they can be appropriately trained, they may be useful for searching through rubble for survivors.
METHOD

Subjects

Ten young-adult pouched rats obtained from APOPO's breeding colony, 4 males and 6 females, served as subjects. They were housed either alone or in pairs and fed a diet of fruit, vegetables, and commercial rodent chow adequate to ensure normal weight gain as compared to controls, although they were mildly food deprived at the time of each testing and training session. From five to seven weeks of age, each rat was handled extensively and exposed to a wide range of stimuli. Details of animal maintenance and socialization are provided elsewhere (Poling et al., 2010b, 2010c). Five randomly-selected rats served as experimental subjects for this study and five others served as controls.

Materials

Throughout the study, to simulate wearing a camera, each rat in the experimental group wore on its back a small nylon "jacket" fastened with a small pouch containing a sound device (see Appendix E). The auditory stimulus used as a return command for the experimental subjects was a slow intermittent beep provided by the sound device and activated by a hand-held remote control. Rats in both groups received mashed banana through a 20 ml syringe presented by trainers. Experimental rats were trained on a 0.5 x 2 m table initially, followed by a 1 x 3 m table, and later in an enclosed area outdoors and in small rooms within our facility. A wide variety of objects (e.g., cardboard boxes, clothing, furniture pieces) were used as obstacles in testing and training. Control rats were trained in specially prepared operant conditioning chambers, which are described elsewhere (Poling et al., 2010b, 2010c).
Training and treatment of control subjects

The control group was included to evaluate the possibility that socialized pouched rats released in a relatively small area would reliably contact humans without being explicitly trained to do so. The five control rats were trained to detect 2,4,6-trinitrotoluene (TNT), a common ingredient in landmines, using our standard training procedures, described elsewhere (Poling et al., 2010b, 2010c). In brief, every weekday a trainer worked with each of these animals to teach through a progressive series of training stages responses culminating with sniffing sample pots containing soil and indicating on any pot containing TNT by holding its nose above the pot for at least 3 seconds. Such responses were rewarded with mashed banana hand-delivered by a trainer. Thus, rats in the control group had experience approaching and receiving food from trainers, but were not specifically trained to search for people. Rats in the control group were targeted for further training and eventual use as accredited landmine detection animals in Mozambique or Angola.

Training and treatment of the experimental subjects

Subjects in the experimental group began training at staggered intervals as age varied by a few weeks. Training comprised five distinct stages, in six different locations. The individuals that served as human targets during training rotated each day and, although several participated more than once, none participated on consecutive days. The human targets consisted of 18 men and 8 women, ranging from 24 to 50 years of age, with 21 from the Republic of Tanzania, 1 from Kenya, 1 from the United Kingdom, 1 from Australia, and 2 from the United States of America. The entire training process took on average 63 days (range: 41 to 80 days).
Conditioning the auditory stimulus as a conditioned reinforcer

Subjects initially received one week of training designed to establish a beep (later to be used as a "return command") as a conditioned reinforcer and as a discriminative stimulus for moving to the location from which the beep originated. In essence, as the rat approached a target person (other than the trainer) while walking around on the ground or table top, the trainer initiated the intermittent beep, and delivered mashed banana to the rat as a reward. This training continued until the rat a) appeared comfortable in the jacket (i.e., no struggling or freezing behavior), b) displayed no fear of novel people, and c) responded to the beep for 10 consecutive trials by turning immediately in the direction of the sound (which was always presented behind the rat), moving to the trainer, and consuming the banana that was presented. Throughout the study, a single training (or test) session typically was arranged for each rat every day from Monday through Friday, at approximately the same time each day (10:00-11:30 a.m.).

Training the search-and-return response

The second stage of training was designed to establish the search-and-return sequence of responses. In this stage, the trainer sat at one end of a 0.5 x 2 m table and the human target sat within 0.5 m of the trainer with his or her arms folded on the table top. For each trial, the rat, wearing its jacket, was released at one end of the table. Once the rat made physical contact with the target, the trainer sounded the intermittent beep for 10 s and, if the rat returned to the trainer within that time, presented a 20 ml syringe of mashed banana to the rat. If two minutes elapsed without contact with the target, or if contact was made but the rat did not respond to the beep within 10 s, no food was presented, and the rat was returned to the start location to begin the next trial. Each time
the rat correctly completed five consecutive trials, the target moved 0.5 m further from
the start location. A researcher using a stopwatch recorded the number of seconds
elapsed before the rat contacted the target, the location of the target, and whether food
was delivered (indicating a successful return to the start position). Each training session
ended after 20 trials to avoid satiation. Throughout this and subsequent training stages, a
correct response sequence was defined as "physical contact with the human target within
30 s of release, and return to the start position within 10 s." For a rat to pass from one
training stage to the next, it had to emit a correct response sequence on at least 85% of
the trials during four consecutive training sessions.

*Training "wait for the return command"

Once a rat completed the first training stage, it was trained to "wait" for the return
command. In this stage, each time the rat contacted the target, the trainer slowly counted
to 3 before sounding the beep. In addition, training moved to a 1 x 3 m table, and the
target changed position each trial. Food was not delivered on trials in which the rat
contacted the target and returned before onset of the beep, but was delivered if it returned
after the onset and before 10 s elapsed.

*Training in multiple locations

In order to disassociate the task from the specific locations used early in training,
the third training stage involved training as just described except the trials were divided
across three different locations: the 1 x 3 m table, a 1 x 1.5 m room, and a 1.5 x 2 m
enclosed outdoor area. The entire training session was reduced to 15 trials with 5 trials
conducted in each location. All three locations were used each day, but the order
changed from day to day. Location, time taken to contact the target, and the number of successful returns were recorded as in prior training stages.

*Training search-and-return with obstacles*

Previous training stages used open areas without any surface-level obstacles (except for some tall thin grass in the case of the outdoor area). In this stage, training moved to a 4 x 4.5 m enclosed indoor room, and obstacles (e.g., cardboard boxes, clay pots, aluminum chairs, plastic containers, and clothing) were arranged on the floor in various positions. Along one wall, a space (1 x 4.5 m) was blocked off using a wooden barrier to separate the search area from the trainer and data recorder. The trainer sat on one side of the barrier, and leaned over as appropriate to either release the rat (from the start location) or to deliver food rewards. Throughout the session, the human target regularly changed physical positions by standing for approximately one-third of the 15 trials, sitting for one-third, and lying down for one-third. Targets did not move, speak, or interact with the rats in any way. For three of the 15 trials, no human target was present in order to train the rat to respond immediately to the return command regardless of whether it had contacted a target. For these trials, a correct (and reinforced) response involved returning to the trainer within 10 s of the beep, which was sounded within 15-60 s of trial onset. A researcher recorded the time taken to contact the target, the return time to the start position, the position of the human target, and the delivery of food. The position of the human target was one of four locations (or absent): on top of a table (with objects arranged to allow the rat to climb up), in a corner completely closed off by obstacles save for a small hole, on the floor surrounded loosely by obstacles, or behind debris that the rat had to climb over to contact the target. After each training day, the
floor was mopped with soap (Dettol Antiseptic Disinfectant; Reckitt Benckiser, Dubai, U.A.E.) and water, and the obstacles were re-arranged. Training was completed for all rats before testing was begun for any rat.

**Testing Procedures**

A 5 x 6 m room, not used during training, was used for the test. The room was arranged similarly to the room used in the last training stage, in that obstacles (e.g., cardboard boxes, aluminum chairs, pieces of clothing) were randomly distributed around the test room and a portion of the room was blocked off by a wooden barrier. The trainer sat behind the barrier and leaned over to either release the rat or deliver food rewards upon successful returns. The target was positioned in one of four pre-determined locations (similar to the last training stage), with the same location used for all rats on a given trial. Each day, the approximate distances from the release point to the four targets were 2, 4, 5.5, and 6.5 m. All 10 rats were treated in the same manner, except that the control rats did not wear a jacket.

The first test session comprised 10 trials. Each trial began by releasing the rat into the enclosed area from the starting location. Once the rat contacted the human target, the trainer sounded the return command, allowing each rat 15 s to return to the starting point to receive the food reward. If three minutes elapsed without contact, the rat was returned to the starting location, no food was delivered, and the next trial began. The same occurred if the rat did not return to the starting point within 15 s of the onset of the return command. A researcher recorded the time to contact the target, the time to return to the starting location, the position of the target, and whether food was delivered. The rats were tested in random order, and the person who recorded data was blind to the
training history of the rat. One day after the first test session ended, a second test session comparable to the first was arranged. Thus, data for 20 trials were available for each rat.
RESULTS

Training results

Figure 9 shows the number of days taken to reach the learning criteria for each rat across all stages. As shown in the figure, the range in the number of days taken to complete a stage was 2 to 25. The length of time to complete each stage varied for each rat, but there was no stage that took longer than 25 days (5 training days per week) to reach the learning criteria.

Test results

Figure 10 shows the percentage of trials in which each rat located the human target within 30 s. The five rats in the experimental group located the person within 30 s in 90 of 100 trials, whereas the five rats in the control group did so on 13 of 100 trials.
Level of detection for individual rats in the experimental and control groups ranged from 65% to 100% and from 0 to 35%, respectively. The overall detection rate of the experimental group was 90% and that of the control group was 13%. To allow for a statistical analysis of these data, rats in the two groups were randomly paired and performance of the five resulting pairs was compared on a trial-by-trial basis. Rats in the experimental group, but not in the control group, located the person within 30 s during 77 of 100 trials. Rats in both groups located the person on 12 of 100 trials, and failed to do so on 1 of 100 trials. One trial occurred in which a control rat, but not an experimental rat, located the target person within 30 s. The ratio of trials in which only an experimental rat located the person relative to trials in which only a control rat located the person was 77 to 1, yielding a sign test value that is significant at the .01 level.

Overall, rats in the experimental group returned to the release location within 15 s of the onset of the return command (beep) on 99 of 100 trials (99%), whereas rats in the control group did not return to the start location on a single trial. Figure 11 shows the average time taken for rats in each group to return to the release point following onset of the return command.
Figure 10. Percentage of trials in which rats in the experimental and control groups located human targets within 30 s.

Figure 11. Time (s) elapsed from the onset of the return command to the rat’s return to the release point. Note that control rats never returned within the specified maximum time of 15 s.
DISCUSSION

The present data suggest that Cricetomys can be trained to locate humans behind or underneath simple obstacles and to return to a release point upon hearing a signal to do so. Rats in the experimental group, which were given explicit training with food rewards, did both reliably, whereas rats in the control group, which had exposure to humans and food rewards but no explicit training to locate people or return on command, seldom if ever did either. This difference illustrates the important of specific training in producing the two responses of interest.

Prior studies have demonstrated that pouched rats can be trained to detect the scent of landmines and human tuberculosis (e.g., Mahoney et al., 2011, 2012; Poling et al., 2010a, 2010b, 2011) and that laboratory rats (Rattus norvegicus) can be trained to detect a variety of visual and olfactory stimuli, including the scent of contraband substances (Otto, Brown, & Long, 2002). Therefore, it is not surprising that pouched rats were successfully trained to find humans in the present study. Although we did not systematically investigate whether the rats used odor or visual cues to find people, observation of the animals suggested that odor cues were important, because the rats often "trailed" the target person if they crossed her or his path.

It is also unsurprising that the sound was readily established as a discriminative stimulus for returning to a release site, because prior research has demonstrated that landmine-detection rats can be readily trained to approach a trainer upon hearing a click when such responding is systematically reinforced (Poling et al., 2000b, 2000c, 2011). The interesting finding of the present study is that two repertoires, searching for a human
target and returning upon command, could be readily established through operant conditioning. These repertoires are necessary if pouched rats are to be used to search for survivors of building collapses, as is our hope.

Of course, training and testing in the present study occurred under relatively simple contrived conditions. Under those conditions, even rats not specifically trained to find people occasionally did so, although less reliably than trained animals. If pouched rats are to be useful for locating survivors of actual building collapses, they will have to perform well under far more challenging circumstances than those arranged in the present study. We are proceeding to train and test the experimental rats used in the present study and other Cricetomys under conditions that are progressively closer approximations to those likely to be encountered in actual search and rescue operations.
GENERAL DISCUSSION

The behavioral approach used in the present project has previously been used to good advantage for measuring, establishing, and improving desired behaviors in a wide range of settings (e.g., Johnson et al., 2001). Therefore, it is not surprising that the approach proved useful reaching the goals of the present project, which were to improve the effectiveness of ongoing operations within an NGO and explore novel humanitarian applications of the organization’s scent-detection rats. In the first experiment, a package intervention was developed to improve the performance of staff in evaluating and conducting animal training sessions. Results suggested that both supervisor and staff performance improved as a result of the package intervention. Moreover, the intervention appeared to be sustainable and cost-effective. In the second and third experiments, two new applications of the scent-detection rats were systematically evaluated. Results of these two studies provide proof-of-principle with respect to the rats’ ability to detect *Salmonella* and to find people. APOPO’s decision makers now possess reliable information that they can use in determining whether to continue either line of research, with the goal of eventually developing an operational system for deploying search and rescue rats or *Salmonella*-detection rats.

A noteworthy aspect of the present project is that the tactics and strategies characteristic of behavior analysis were used to improve organizational efficiency and to investigate opportunities for growth, both of which are essential components for organizational sustainability (Weerawardena, McDonald & Mort, 2010). Importantly, this approach had three characteristics that made it well suited for use in an NGO.
working in resource-poor areas. First, the investigations were conducted in a cost-effective manner, requiring only regular staff members, one external consultant, and a few materials for data collection. In the first experiment, all of the supervisors readily learned how to use the job aid and demonstrated high performance scores without any assistance from the consultant. These data suggest that long-term benefits can be obtained with little to no assistance from external experts. Second, each experiment involved a direct measure of behavior change (whether human or rat) and visual analysis of findings, rather than indirect measures of performance (e.g., responses to questionnaires) and statistical analyses, which increases the face validity of obtained results and makes them easier for local staff to understand. Third, each study involved a relatively small number of participants, which reduces both the time and expense required to complete it. A major strength of the behavior-analytic approach to research is that it makes good use of resources (Poling, Methot, & LeSage, 1995), and this strength was clearly evident in the present project.

Of course, each study was designed to be a simple demonstration of the potential value of behavioral techniques and did not fully investigate the topic of concern. For instance, a limitation of Experiment 1 is that follow-up data demonstrating long-term improvements and sustainability were not collected. Similarly, determining whether rats can be used operationally to detect *Salmonella* bacteria in the feces of infected horses (or people), or find people trapped under rubble, will require additional research. The data reported in this manuscript were sufficiently promising that APOPO is pursuing such research.
Although challenging, the present project was successfully completed and
demonstrates that the same general strategies used to benefit other kinds of organizations
can be of value to NGOs operating in resource-poor areas. One factor that almost
certainly contributed to the project’s success is that the behavioral consultant lived in
Tanzania throughout the project. Maintaining intervention integrity – that is, insuring
that conditions are arranged and data collected as intended – is a challenge in any
research setting and this challenge is magnified when working in resource-poor and
multi-cultural settings. By being on-site, the behavioral consultant could monitor
ongoing activities on a daily basis and, in most cases, obviate or rapidly correct problems.
Doing so became easier over time, as she came to better understand organizational
culture of APOPO and the Tanzanian culture in which APOPO is embedded. Success
anywhere requires cultural sensitivity, which is more easily desired than attained.
FUNDING

This research was supported in part by the Society for the Advancement of Behavior Analysis 2011 International Development Grant. A portion of this work was based upon work supported by the U.S. Army Research Laboratory and the U.S. Army Research Office under grant number W911NF-11-1-0464.
REFERENCES


in institutionalism in organizational analysis. Chicago, IL: University of Chicago Press.


Appendix A

Human Subjects Institutional Review Board Approval Letter
Western Michigan University

Date: October 9, 2012
To: Alan Poling, Principal Investigator
    Amy Durgin, Student Investigator
From: Amy Naugle, Ph.D., Chair
Re: HSIRB Project Number 12-10-18

This letter will serve as confirmation that your research project titled “Task Clarification and Performance Feedback: Procedures for Improving and Maintaining Supervisor Performance in a Non-Governmental Organization in East Africa” has been approved under the expedited category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may only be conducted exactly in the form it was approved. You must seek specific board approval for any changes in this project (e.g., you must request a post approval change to enroll subjects beyond the number stated in your application under “Number of subjects you want to complete the study”). Failure to obtain approval for changes will result in a protocol deviation. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

Reapproval of the project is required if it extends beyond the termination date stated below.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: October 9, 2013
Appendix B

The Job Aid
SOIL FLOOR TRAINING

1) If the Supervisor is uncertain, ask the trainer what he is trying to teach the rat today □

2) Check the box that best answers this question and then read the list of statements below.
   a. Circle “Yes” if the statement is true, or “No” if the statement is false. If at any point the answer could be "No" - circle "No".
   b. If the trainer is not doing something correctly, instruct the trainer to do the item the correct way, and mark the box next to “FB” to show that you have given the trainer “Feedback”.

What is the trainer working on with this rat? (more than one box can be checked at a time, except for the last)

- [ ] 1-Training the rat to scratch on positive tea-eggs.
- [ ] 2-Training the rat to discriminate between positive and negative tea-eggs.
- [ ] 3-Shaping the rat to respond to the click, by walking immediately and completely to the trainer (both sides)
- [ ] 4-Training the rat to walk in one direction and to finish lanes.
- [ ] 5-Trainer never steps into training box
- [ ] 6-Shaping strong indications on positive tea-eggs.

1. Type and position of Tea-Eggs is appropriate
   - [ ] If training #1: All or majority of tea-eggs are positive
   - [ ] If training #2-6: positive and negative tea-eggs are used
   - [ ] Buried at appropriate depth?
     - [ ] Rat 1st day = on top
     - [ ] Rat 2+ days of training = slightly covered
     - [ ] Rat learning strong indications = buried deeper

2. The rat appears comfortable
   - [ ] The rat is sniffing and exploring the area freely; the rat is NOT scared or stiff

3. The rat appears motivated
   - [ ] After sniffing a positive tea-egg, the rat immediately scratches/digs/bites
   - [ ] After hearing the click, the rat responds immediately by moving toward the trainer (or looking around as a reaction to the click)
   - [ ] The # of responses to the click are greater than the # of no responses

4. Trainer never pulls the rope/rat
   - [ ] Short tugs are allowed when the rat has stopped to groom for greater than 15 seconds

5. Timing of the click is appropriate
   - [ ] When tea-eggs are on the surface - the click is timed with sniffing/contact with the tea-egg (not when rat is biting the tea-egg)
   - [ ] When the tea-egg is buried, the rat scratches and the trainer clicks quickly

6. Trainer never steps into training box
   - [ ] If training #3 - the trainer steps into the training box and
gradually moves toward the side to train the rat to walk to the side of the box

<table>
<thead>
<tr>
<th>7. The trainer practices on both sides of the box</th>
<th>Y / N</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>• <strong>If</strong> the rat has already learned to walk to one side of the training box for 5+ consecutive trials, the trainer moves to the opposite side to train the rat on both sides of the box</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8. Trainer follows the &quot;one direction&quot; rule</th>
<th>Y / N</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Two trainers are training</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>• The rat walks in one direction across the lane.</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>• If the rat attempts to walk in the opposite direction, the trainer holds the rope tight (no pulling!)</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. The correct trainer clicks and rewards the rat</th>
<th>Y / N</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The trainer rewards the rat for responding to the person who clicked and does not give food if the rat returns to the wrong trainer (who did not click).</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10. Measuring strength of indications</th>
<th>Y / N</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>• If the trainer is shaping strong indications, the trainer is counting out loud or using a stopwatch</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11. The trainer is focused on training the right skill for this rat</th>
<th>Y / N</th>
<th>FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The rat demonstrates the skills listed above the skill being trained today</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>
Appendix C

Treatment Acceptability Rating Form-Revised
1. How clear was your understanding of how to use the job aid?

Not at all (2) Neutral (3) Very clear (5)
Clear (1)

2. How acceptable was the job aid as a tool for evaluating the trainer's performance?

Not at all (2) Neutral (3) Very acceptable (5)
Acceptable (1)

3. How willing would you be to use the job aid on a regular basis?

Not at all (2) Neutral (3) Very willing (5)
Willing (1)

4. How much did you *like* using the job aid?

Not at all (2) Neutral (3) Liked using it very much (5)
Like (1)

5. How well would using the job aid on a regular basis fit into your supervisory routine?

Not at all (2) Neutral (3) Very well (5)
Well (1)
Appendix D

Diagram of the Semi Automated Cage
Appendix E

Photos of Training the Search Rats
Trainer feeding the rat mashed banana and pellet. Notice the blue "jacket" adorned by the rat. The trainer uses a cell phone (shown in left hand in the picture) to initiate the return command, which is then produced by a small device inside the jacket.

In this picture, the rat is about to make contact with the trainer, who watches casually over his shoulder.