

Evaluation of the Validity of a Maximum Likelihood Adaptive Staircase Procedure for Measurement of Olfactory Detection Threshold in Mice

Amy C. Clevenger and Diego Restrepo

Department of Cell and Developmental Biology, Neuroscience Program, and Rocky Mountain Taste and Smell Center, University of Colorado School of Medicine, University of Colorado at Denver and Health Sciences Center at Fitzsimons, Mail Stop 8108, PO Box 6511, Aurora, CO 80045, USA

Correspondence to be sent to: Amy C. Clevenger, Department of Cell and Developmental Biology, Neuroscience Program, and Rocky Mountain Taste and Smell Center, University of Colorado School of Medicine, University of Colorado at Denver and Health Sciences Center at Fitzsimons, Mail Stop 8108, PO Box 6511, Aurora, CO 80045, USA. e-mail: amy.clevenger@uchsc.edu

Abstract

Threshold is defined as the stimulus intensity necessary for a subject to reach a specified percent correct on a detection test. MLPEST (maximum likelihood parameter estimation by sequential testing) is a method that is able to determine threshold accurately and more rapidly than many other methods. Originally developed for human auditory and visual tasks, it has been adapted for human olfactory and gustatory tests. In order to utilize this technique for olfactory testing in mice, we have adapted MLPEST methodology for use with computerized olfactometry as a tool to estimate odor detection thresholds. Here we present Monte Carlo simulations and operant conditioning data that demonstrate the potential utility of this technique in mice, we explore the ramifications of altering MLPEST test parameters on performance, and we discuss the advantages and disadvantages of using MLPEST compared to other methods for the estimation of thresholds in rodents. Using MLPEST, we find that olfactory detection thresholds in mice deficient for the cyclic nucleotide-gated channel subunit A2 are similar to those of wild-type animals for odorants the knockout animals are able to detect.

Key words: CNGA2, knockout mice, maximum likelihood, odor detection, olfactometer, operant conditioning

Introduction

Olfaction is a complex sensory modality in which neurons responsible for transmitting information to the olfactory bulb (OB) express a single olfactory receptor subtype and mediate detection of a subset of structurally related odorants (Buck, 2000; Xu *et al.*, 2000). In the OB, these neurons synapse in glomeruli to create distinct spatiotemporal activity patterns that are thought to be the basis for an animal's ability to detect small concentration differences of a single odorant and discriminate between odorant molecules that differ in as few as a single carbon atom or even the location of that carbon atom. The use of gene-targeted mice has become increasingly necessary to delineate the complexities of this system. Phenotypic changes produced by specific alterations in the genetic makeup of an animal may be subtle, however, requiring the concomitant development of sensitive, straightforward experimental paradigms that will be able to detect small behavioral changes.

Cyclic nucleotide-gated channel subunit A2 knockout (CNGA2 KO) mice are one example in which the accurate

assessment of olfactory ability is critical. In these mice, targeted disruption of the CNGA2 prevents signal transduction through the adenosine 3',5'-cyclic monophosphate (cAMP) second messenger pathway (Brunet *et al.*, 1996; Baker *et al.*, 1999; Lin *et al.*, 2004). Initial experiments utilizing electro-olfactogram technology, which measures odor-induced field potential changes at the surface of the olfactory epithelium (OE), found the strain unable to respond to all 12 odorants tested, suggesting that the cAMP-mediated second messenger pathway is the only signal transduction mechanism present in the main olfactory system of mice (Brunet *et al.*, 1996). Continued characterization of these animals using an automated olfactometer system developed by Bodyak and Slotnick (1999), however, resulted in the finding that CNGA2 KO animals are able to smell some odorants, including ethyl acetate (EA) (Lin *et al.*, 2004). This observation, taken together with studies of the responsiveness of the OE and OB to these odors in CNGA2 KO mice, provides evidence for the existence of alternate signal transduction mechanisms

in the mouse main olfactory system. Further characterization of the olfactory ability of CNGA2 KO animals could provide interesting insights into olfactory processing. For example, for those odorants CNGA2 KO animals can detect, do they differ from wild-type (WT) controls in terms of detection threshold? Given the responsiveness of CNGA2 KO to a variety of odorants, we need a fast and accurate measure of olfactory detection that will allow us to determine threshold to several different odorants in these mice.

Sensory detection threshold can be defined as the stimulus intensity necessary for a subject to perform at criterion (i.e., percent correct response) on a detection test (Krantz, 1969; Harvey, 1986; Wichmann and Hill, 2001). The majority of odor detection threshold determinations in rodents have been performed using the descending method of limits (Pietras and Moulton, 1974; Slotnick and Schoonover, 1993; Apfelbach *et al.*, 1998; Youngentob and Margolis, 1999; Vedin *et al.*, 2004; Youngentob *et al.*, 2004; Pho *et al.*, 2005), a method where animals are asked to detect successively lower odor concentrations until they reach a concentration that they cannot detect. The last concentration detected by the mouse is considered to be the estimated threshold. The descending method of limits has the limitation that the estimated threshold must be one of the tested concentrations and that the continued exposure to subsequently lower odor concentrations provides an opportunity for the animal to cue on nonchemosensory signals. Because of these drawbacks, we sought to use a different method for odor threshold determination in mice.

Maximum likelihood parameter estimation by sequential testing (MLPEST) (Harvey, 1986, 1997; Linschoten *et al.*, 2001) is a method that can be utilized to determine olfactory detection threshold using fewer trials and resulting in a more accurate determination of threshold values than methods such as the one-up-two-down or other related variants of the staircase method (Wetherhill and Levitt, 1965) and the ascending method of limits (Cain *et al.*, 1983; Apter *et al.*, 1999). Because most detection threshold data approximate a Weibull psychometric function (Linschoten *et al.*, 2001), MLPEST methodology allows rapid threshold determination by fitting the experimental data to this function at the end of each trial and choosing the next stimulus concentration in a manner that allows rapid convergence to a threshold. In human experiments this was found to give accurate taste and smell threshold measurements (Linschoten *et al.*, 2001).

In order to determine olfactory detection threshold in CNGA2 KO mice, we modified MLPEST methodology to work with an olfactometer system that uses a go-no-go paradigm (Bodyak and Slotnick, 1999) and explored the validity of the use of this method in mice. We found that MLPEST yielded quick and reproducible determination of detection threshold. Interestingly, however, thresholds determined by MLPEST were several orders of magnitude higher than the thresholds determined by the descending method of limits. The reason for the differences in thresholds determined with these two methods is debatable. On the one hand,

the difference in threshold could be due to a decrease in the detection threshold due to improved sensitivity of the animal to the odor caused by the more substantial training during the descending method of limits. Such improvement with training has been demonstrated in the visual, auditory, and somatosensory systems, where it has been named "perceptual learning" (Lu and Doshier, 2004; Witte and Kipke, 2005). On the other hand, the difference in thresholds could be due to inaccurate determination of thresholds by the MLPEST method due to insufficient training to attain a suitable attentional status of the animal. Without further experimentation, we cannot distinguish between these two possibilities. Our experiments indicate that MLPEST is a promising method for threshold determination and that the reason for the discrepancy in threshold values determined with these two methods should be tested more thoroughly in the future.

Using MLPEST we determined detection thresholds in CNGA2 KO mice and controls. We did not find a difference in MLPEST detection threshold between CNGA2 KO mice and their WT littermates.

Materials and methods

Animals

Adult male CNGA2 KO mice, their WT littermates, and FVB mice were used. CNGA2 KO mice were generated by crossing heterozygous CNGA2 KO female mice (Brunet *et al.*, 1996; Baker *et al.*, 1999) (kindly provided by Drs John Ngai, University of California, Berkeley, CA, and Randall Reed, Johns Hopkins, Baltimore, MD) with WT 129/SvJ male mice (Jax Mice, Bar Harbor, ME). This breeding protocol results in the generation of CNGA2 KO hemizygous mice because the CNGA2 gene is located in the X chromosome. Genotyping was completed by amplification of genomic DNA with the polymerase chain reaction (Lin *et al.*, 2004). Prior to testing, mice were housed in groups of no more than five animals per cage and given food and water *ad libitum*. During testing, mice were housed singly and given food *ad libitum*. All mice were maintained in a 12:12 h light:dark cycle prior to and concurrent with testing. During water restriction, animals were kept between 80% and 85% of their unrestricted baseline weight. All procedures were performed under protocols approved by the animal care and use committee of the University of Colorado at Denver and Health Sciences Center.

Odorant preparation

We followed procedures similar to those of Bodyak and Slotnick (1999). Briefly, high-purity odorants were obtained from Aldrich Chemical Company (Milwaukee, WI), Fluka (Ronkonkoma, NY), or Takasago Corporation (Shinagawa, Japan). Odorant bottles were obtained from Qorpak (Bridgeville, PA). Bottles were washed in 95% ethanol and placed in a dedicated oven to dry. Odorless tubing (Cole

Palmer, cat. no. 06424-67; Vernon Hills, IL) was placed through holes drilled in the caps for these bottles. Prior to each experiment, odorant bottles were filled with 10 ml of odorless mineral oil (MO) and “blanked.” Blanking consisted of trained mice comparing two bottles and not detecting a difference for 10 blocks (200 trials). Odorant bottles were then prepared by diluting odorant in this MO. Odorant concentrations are stated as percent dilution (v/v) in MO. Each concentration was made fresh on the first day of MLPEST testing and used for a maximum of 2 weeks. Except at high liquid odor concentrations, use of liquid dilution results in air odorant concentrations that are proportional to the odor concentrations in the MO (Cometto-Muniz *et al.*, 2003).

Behavioral training

The olfactometer setup and training have been described previously (Bodyak and Slotnick, 1999) (Knosys Olfactometers Inc., Tampa, FL, <http://knosysknosys.com>). Briefly, mice were deprived of water for 2 days. On the third day, mice were placed in the olfactometer chamber, during which they were able to trigger the beginning of a trial by poking their nose into the odor delivery chamber and interrupting a photobeam. In an initial session, the mice were trained to lick a water delivery port present inside the odor delivery chamber. During the second day, mice were trained to respond to an S+ odor rewarded with water delivery. A trial consisted of stimulus presentation through the odor delivery port for 2 s. The odor stimulus was presented during each trial after a variable time period (1–1.5 s) during which odor delivery to the odor delivery chamber was bypassed to an exhaust tube through the actuation of a “final valve.” The use of the final valve resulted in abrupt onset of the odor stimulus at the beginning of stimulus presentation. For stimulus presentation, the headspace of one of eight odorant bottles was further diluted 40 times with clean air before being introduced to the odor delivery chamber. Odor stimuli were characterized as S+ (rewarded stimulus, 1% v/v EA diluted in MO) or S– (unrewarded stimulus, MO alone). On an S+ trial, mice that licked at the water delivery port in each half second of the trial period received a 5- μ l water reward. On an S– trial, mice were given no water reward and eventually refrained from licking at the water delivery port. A correct response was licking on an S+ trial or the absence of licking on an S– trial. After 20 trials, the percent correct for that block was determined. Following completion of three consecutive blocks with a score greater than or equal to 85%, training was considered complete. Typically, a well-motivated mouse achieved criterion in two sessions of S+/S– training. However, some mice required further water deprivation and retraining.

Threshold determination by the descending method of limits

FVB mice were tested in go-no-go sessions similar to the second day S+/S– session described above (see Behavioral

Training). The S+ stimulus was EA, and the S– stimulus was MO. Subsequently, lower EA odor concentrations ranging from $10^{-2}\%$ to $10^{-6}\%$ were tested in each of the go-no-go sessions. Each session was terminated when a mouse reached criterion (when it responded 85% correct or better on three consecutive blocks of 20 trials). Whenever possible, two sessions were run in each day. Attempts to run more than two sessions per day were unsuccessful because the mice would not behave in the olfactometer in the third session. Testing was stopped when the mouse responded below 85% in six blocks. A session where the mouse was asked to detect the difference between two vials containing MO (no EA added) was run to ensure that the mice were not cueing on nonchemosensory stimuli.

Programs for control of odor detection trials in the Bodyak and Slotnick olfactometer

The olfactometer was controlled using a set of programs written in C. These programs are available at <http://www.uchsc.edu/rmtsc/restrepo> (under Biomedical Info and Tools). Two of the programs (begin.exe and splussminus.exe) perform the training described in the previous section. These programs perform the same sequence of signal detection events and valve openings attained by the BASIC programs provided with the olfactometer of Bodyak and Slotnick (1999). We rewrote the programs in C because this language allows for more efficient control of the central processing unit and because the code used to write the MLPEST program was kindly provided to us (by Dr Lewis O. Harvey, University of Colorado, Boulder, CO) as routines written in C (Harvey, 1997). Use of the same routines written in the same computer language in the training programs and the MLPEST programs ensures that the timing of events in each trial is exactly the same in the training and MLPEST tests.

Threshold determination by MLPEST

The day after training was complete mice were run in the MLPEST program. Throughout the experiments, it was important to ensure that animals were under stimulus control—responding to the rewarded (S+) odorant. To ensure that the animal began MLPEST under stimulus control, the first block (20 trials) consisted of the same S+/S– paradigm used in training (1% odorant in MO vs. MO alone). If the mouse scored lower than 85%, the S+/S– block was repeated. If the mouse scored 85% or higher, subsequent blocks of 30 trials were presented. In these blocks, odorants were given in groups of three: one S+, one S–, and one test concentration (given in random order). In Results, we refer to the trial performed with the test concentration as the “test trial.” The responses in the test trials are used by MLPEST to estimate the maximum likelihood threshold value. The S+ odorant was 1% odorant in MO. Test concentrations were diluted in log steps, from 1 to $10^{-6}\%$ odorant in MO (throughout the article percent dilution is understood to

be v/v). The S− odorant was always pure MO. Test concentrations were rewarded randomly at a rate of 50%. In order to further strengthen stimulus control, mice were given four “refresher” S+/S− trials (without interspersed test trials) following any false alarm response during the MLPEST trials. At the end of each block of 30 trials, the percent correct responses on S+ and S− were computed to demonstrate the maintenance of stimulus control. When animals performed below 85% for two consecutive blocks or below 80% for any block during an MLPEST session, data for the entire session were discarded.

MLPEST is an automated program that finds the threshold value that is most likely to represent the experimental data when fitted by the Weibull psychometric function (Figure 1; Maloney, 1990; Linschoten *et al.*, 2001; Wichmann and Hill, 2001). This asymmetric function has four variables.

1. γ is the intercept on the vertical axis (the axis for percent correct). In the olfactometer, γ indicates the probability of a response at undetectable odorant concentrations (the false alarm rate), which varies around a value of 10% for mice in this study (see Results section under “Psychometric function and false alarm rate”). The value of γ utilized by MLPEST is the running average of γ determined from all the S− trials within each run. This value is recalculated after each block of 30 trials and therefore changes dynamically as the experiment progresses, minimizing problems due to changes in false alarm rate during the session.
2. β is a parameter reflecting the slope of the curve at its steepest point. In the MLPEST program, we used a value of 3.5 that had previously been used by Linschoten and co-workers (2001). In Results we show that, consistent with earlier reports (Terutwein and Strasburger, 1999;

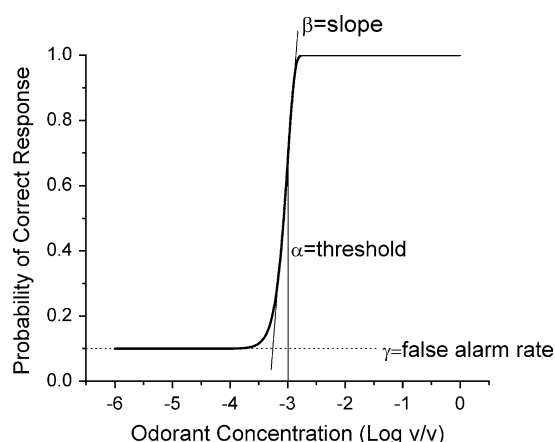


Figure 1 Weibull psychometric function. γ : false alarm rate; β : steepest slope; α : stimulus concentration at the point on the curve with the steepest slope, threshold; δ : lapse rate [difference between 100% and the response at infinitely high odor concentrations (not shown on graph)]. This particular example is for $\alpha = -3$, $\beta = 3.5$, $\delta = 0$, and $\gamma = 0.1$.

Linschoten *et al.*, 2001), determination of threshold by MLPEST is relatively insensitive to the choice of the value of β utilized by the program.

3. α is a scale parameter that indicates the point on the curve with the maximal slope (Harvey, 1997). α is the threshold value that the MLPEST program estimates. For the Weibull function, α lies at 64% correct response rate.
4. δ is the difference between the percent correct response at infinitely high odor concentrations and 100%. This parameter, which is a measure of lapses by the observer at high odor concentrations, was set to zero for our experiments because mice lapse rarely.

The MLPEST procedure is discussed in detail by Harvey (1986, 1997) and Linschoten *et al.* (2001). As modified for the Bodyak and Slotnick olfactometer, MLPEST computes the likelihood for 1000 values of α spaced logarithmically throughout the entire range of stimulus concentrations (1%–10^{−6}% odorant in MO in this particular case). Before each trial, MLPEST calculates the likelihood for each of these values of α based on all the data accumulated in prior trials. The candidate α , representing the current estimate of the sensory threshold, is then assigned to the value of α that has the maximum likelihood (this can be thought of as the value that results in the best fit of the Weibull function to the data).

MLPEST is started with a guessed threshold given by the user. The program then simulates a Gaussian observer with this threshold to populate a second likelihood (the posterior likelihood, which we will refer to as α') that incorporates both the expectations provided by the user in the guessed threshold as well as the data accumulated in prior trials. The α' with the largest likelihood in the posterior likelihood α' array is then used to choose the next stimulus (while α could be used for choosing the next stimulus, α' is used instead because α is highly variable during the first several trials). Thus, at the end of each trial, MLPEST determines three parameters.

1. A new estimate for α that is the maximum likelihood α chosen among the entire array of calculated α values based on the data obtained in all previous trials (but not including the simulated trials for the Gaussian observer).
2. The threshold computed from the posterior likelihood α' (both the Gaussian trials and the previous trials by the animal). In previous taste and smell studies in humans using MLPEST methodology, Linschoten and co-workers (2001) have used this posterior likelihood method to determine the next test concentration for each trial. Throughout the article we use the term “posterior likelihood method” to describe this method for choosing the next stimulus.

3. The stopping confidence interval (CI) for α . When the stopping CI reaches a low value entered by the user, the program terminates, yielding an estimate for the threshold (the α computed in the last trial).

Monte Carlo simulations

We used computer simulations to determine the accuracy of MLPEST-estimated threshold values under constraints imposed by the Bodyak and Slotnick olfactometer. All simulation experiments used the Monte Carlo technique, which is utilized for complex simulations in a wide array of physical and social sciences (Press *et al.*, 1992). Our Monte Carlo simulations followed a parametric bootstrap method that allowed us to determine bias and variability in the estimated threshold values (Maloney, 1990; Linschoten *et al.*, 2001; Wichmann and Hill, 2001). The bias was determined as the difference between the value of the threshold estimated by MLPEST and the true value of the threshold while the variability was estimated with two measures: the standard deviation (SD) and the 68% CI. In the parametric bootstrap method, a Monte Carlo observer samples randomly from the probability density function used to describe the system. For our simulations, a Weibull probability density function was used to model the responses of the simulated mouse, as this function approximates the experimental data closely.

We parsed parameter space for the Weibull function describing the response of the simulated mouse as follows:

1. The threshold α was varied throughout the entire range of concentrations (from 1% to $10^{-6}\%$) in logarithmic steps of 0.5.
2. As shown in Results, the value of $\beta = 0.6$ is a lower limit estimate of β for the psychometric function describing the behavior of individual mice. Most simulation results shown in figures throughout the article are for $\beta = 0.6$. We also performed simulations with $\beta = 3.5$. Any differences for simulations at $\beta = 3.5$ are mentioned in the text. A limited number of simulations were run with β between 3.5 and 0.6 to ensure that parameters changed monotonically as a function of β (not shown).
3. For real mice, γ generally fell between 0 and 0.2. For the simulated mouse, γ was set to 0.1. In addition, a few simulations were run with different values of γ ranging from 0 to 0.2, but we did not find any major difference with the simulations at $\gamma = 0.1$ (data not shown).
4. Because real mice lapsed rarely, we used a value 0 for δ .

After setting specific values for the Weibull parameters for the simulated mouse and choosing a stopping CI (0.5 unless otherwise stated), the simulation program (mlpestsim.exe) simulated 500 MLPEST sessions with Monte Carlo observers ("silicon mice"), sampling from the Weibull distribution. The random number generator was initialized separately for each mouse to ensure independent behavior. For each

MLPEST session, the program saved values of estimated α and γ as well as the total number of trials and number of test trials required to attain the specified stopping CI.

Statistics

Because the distribution of estimated α did not follow a normal distribution (not shown), a Wilcoxon rank sum test or Friedman's analysis of variance (ANOVA) were used to evaluate the statistical significance for differences between estimates of α . Regular ANOVAs were used for all other comparisons. The Wilcoxon test and the ANOVAs were calculated using MATLAB (The MathWorks Inc., Natick, MA). Least-squares fits of the Weibull function were performed using Origin (OriginLab Corporation, Northampton, MA).

Results

Threshold estimation with the descending method of limits

Figure 2 shows the results of threshold determinations using the descending method of limits. The mice were initially trained to differentiate between MO and a 0.01% concentration of EA in MO (hereafter referred to as the logarithm of the concentration: -2). In subsequent sessions, mice were asked to differentiate between MO and increasingly more dilute concentrations of EA in MO (spanning the concentration range from -2.3 to -6). As shown, the responsiveness of the animal in Figure 2a to EA dropped substantially when the odor concentration was dropped from -5 to -6 . The mouse was also tested for discrimination of MO versus MO (Figure 2a) to ensure that it was not responding to non-chemosensory cues. Defining the threshold as the last concentration where the animal reached criterion in six blocks or less resulted in a threshold value for EA for this mouse of -5 (the same threshold value was obtained with another animal, and a threshold of -8 was obtained with a third animal; criterion was defined as performing 85% or better in three consecutive blocks).

We found two potential disadvantages of using the descending method of limits to estimate detection thresholds. 1) Because animals would stop working when more than two sessions were run per day, determination of the threshold took at least 3–4 days. 2) More importantly, of five mice tested with the descending method of limits two (40%) responded to MO (>85% correct in six blocks or less), presumably because these mice were responding to nonchemosensory cues (Figure 2b). Indeed, we could show that some mice that responded to MO were responding to somatosensory cues because cutting their vibrissae brought the response level to chance (not shown). These data suggested that the training procedure necessary for threshold determination using the descending method of limits provided ample chance for the mice to cue on nonchemosensory stimuli. Because of these issues, we sought to develop a method to generate fast

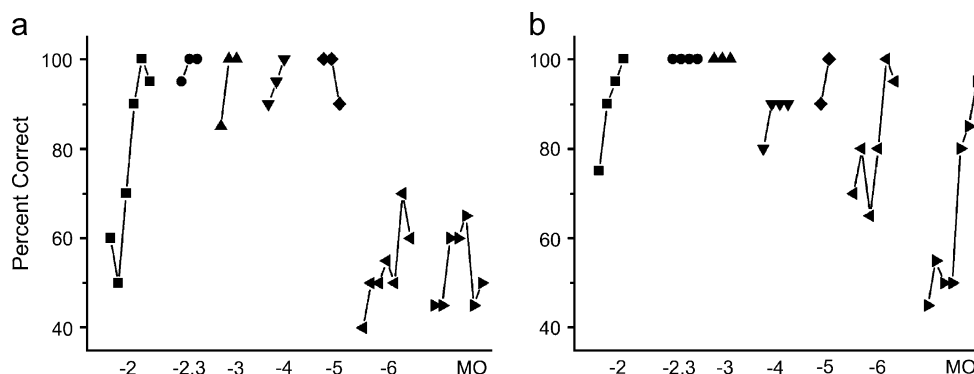


Figure 2 Determination of threshold in two mice by the descending method of limits. The data show percent correct in a detection task as a function of block number and EA concentration. Percent correct detection of EA in MO compared to MO was determined for each block of 20 individual trials. The percent correct for each block is denoted with a symbol, and data points for each block within each session are connected with solid lines. Each session was run at different EA concentrations (diluted in MO). EA concentrations for each session are denoted under the horizontal axis. As a control, to ensure that mice were not responding to nonchemosensory cues, we tested the ability of the mouse to discriminate between two vials containing MO (this session is denoted MO in the horizontal axis). The mouse in **(a)** did not respond in the MO versus MO control, while the mouse in **(b)** responded, indicating it was not under chemosensory stimulus control.

(single session) and accurate measures of detection threshold using MLPEST.

MLPEST: performance of posterior likelihood method

In order to adapt MLPEST to the Bodyak and Slotnick olfactometer, it was important to take into account the fact that mice will only complete approximately 200 trials before water satiety ensues. The olfactometer also limits the number of test concentrations because it only contains enough valves for eight odorant bottles. This limitation could be particularly problematic because psychometric functions for olfactory detection in mice often span several log units of odorant concentration (Youngentob and Margolis, 1999). Simulation experiments were therefore conducted to determine whether MLPEST would be able to calculate α , in less than 200 trials, within a stopping CI of 0.5 (the criterion used by the program to determine when to stop) if test concentrations differed by one log unit. In these experiments, simulated mice were programmed to perform with a γ of 0.1, a β of 0.6, and a δ of 0. These values are similar to the values describing the psychophysical Weibull function for experimental animals (Results section under “Psychometric function and false alarm rate”). The guessed threshold and actual threshold were varied to include all possible combinations. The simulation program completed MLPEST trials for 500 silicon “mice” for each parameter combination.

From the initial simulation experiments, we determined that when using the posterior likelihood method to obtain the next test stimulus, the program yielded estimates of α that were less skewed if a high concentration was used as the initial guessed threshold. For example, when a -6 concentration was used as the guessed threshold, the program made errors which skewed toward the guessed threshold of -6 when the actual threshold was -1.5 (Figure 3a). In contrast, if the guessed threshold (0) was higher than the actual threshold (-4.5), the variance was

not as skewed (Figure 3b). Notice that as expected because the stimuli were all chosen to be whole numbers (-6 to 0), the histogram of threshold values was discontinuous favoring the whole numbers in the logarithmic scale.

The simulations indicated that, for the posterior likelihood method, the guessed threshold should be a high value in order to minimize skewness in the histogram of measured thresholds. Indeed, this worked well when the difference between the guessed threshold and the actual threshold was <4 . Thus, for example, for a guessed threshold of 0 and actual threshold of -4.5 , only 18% of the sessions required more than 200 trials to reach criterion (Figure 3c). However, when the difference between the guessed threshold and the actual threshold was larger than 4, convergence to a threshold value took place after an unacceptably large number of trials. Thus, with a guessed threshold of 0 and an actual threshold of -6 , 81.2% of the mice required more than 200 trials to reach the stopping criterion of 0.5 (Figure 3d). This meant that, under these conditions, 81.2% of the experiments would be prone to fail because the mice would likely stop responding after 200 trials. Convergence was even slower when β for the simulated mouse was increased to 3.5 (data not shown). Since we could not predict the true threshold values for odorants tested with real mice, we sought to develop a different method for the determination of the next test concentration that would result in fewer trials per run.

The randomized window method requires fewer trials at the expense of a slightly larger variability in the estimate of α

All the simulation experiments described in the previous section were run utilizing the posterior likelihood method for choosing the next test stimulus after each trial (see Materials and Methods). In order to decrease the number of trials needed to bring the animals to criterion, a new method to

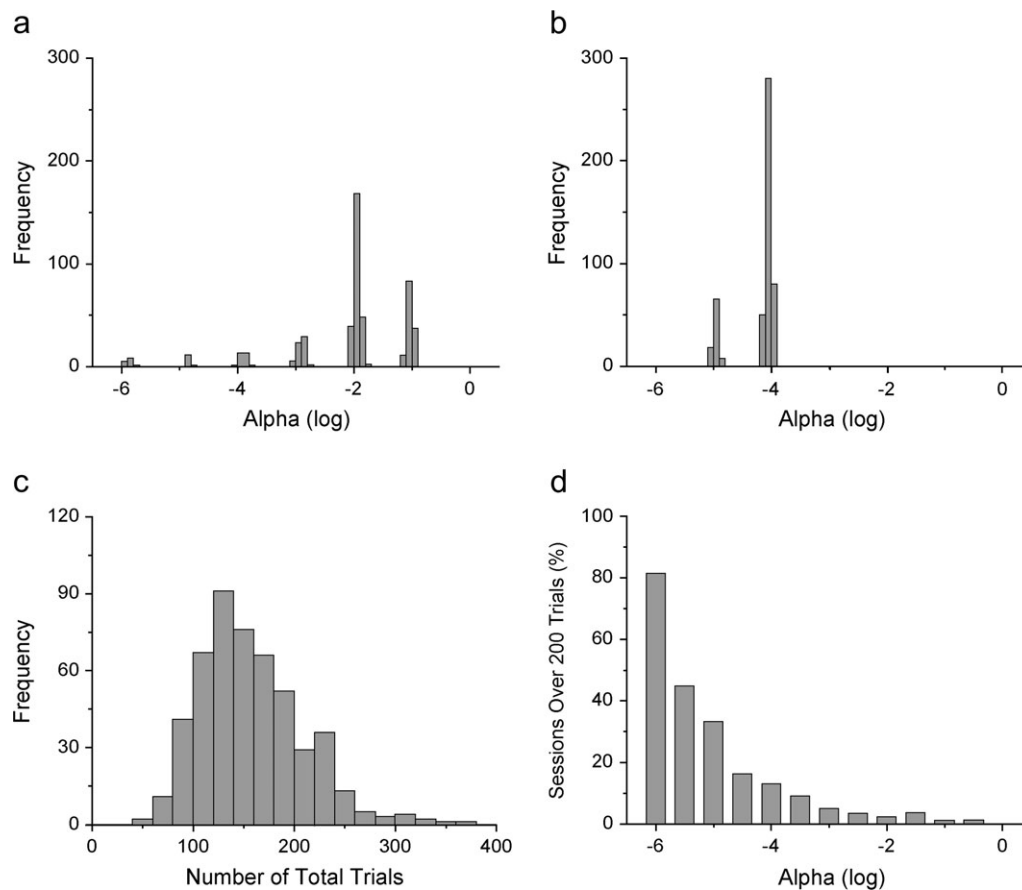


Figure 3 Posterior likelihood method utilized in 500 simulation runs. **(a)** Histogram of α s (thresholds) produced when guessed threshold is $10^{-6}\%$ (shown as -6) and actual threshold is $10^{-1.5}\%$ (-1.5). The mean α was -2.11 , the median -1.94 , and the SD 1.07 . **(b)** Histogram of α s (thresholds) produced when guessed threshold is 1% (0) and actual threshold is $10^{-4.5}\%$. The mean α was -4.22 , the median -4.07 , and the SD 0.36 . **(c)** Histogram of the total number of trials to criterion for a guess α of 0 and actual α of -4.5 . The average number of trials to criterion was 157 , and 18% of the 500 sessions required more than 200 trials to complete. **(d)** Percentage of runs that required over 200 trials to reach criterion when guessed threshold is 1% and actual threshold varies from $10^{-0.5}\%$ to $10^{-6}\%$.

determine the next stimulus concentration was created. We reasoned that in order to optimize convergence to threshold, we needed a method to select subsequent odorant concentrations so that the mice would be presented with approximately the same number of test concentrations above and below the estimated α . We implemented a procedure that we termed the “randomized window method.” In this procedure, after α was calculated at the end of each trial, the program determined how many data points lay on either side of the current estimated α . The program then chose the subsequent test stimulus to lie on the side of α with fewer data points, randomly within a window of two concentrations from the estimated α . In other words, if the current estimate were -3 and more data points were found at higher concentrations, the next test stimulus would be either -3 or -4 . In this way, we reasoned that the randomized window method would be able to choose test stimuli which would modify the estimated α more rapidly than the posterior likelihood method.

Shown in Figure 4 a,b are individual simulations that illustrate the difference in the convergence to threshold between

the posterior likelihood method and the randomized window method. Because the posterior likelihood method always chooses a test stimulus with a probability that favors the value closest to the current estimate of α , many test concentrations in a row will be identical, even when the “mouse” response consistently deviates from the current estimate of α . In contrast, with the randomized window method, the program adapts quickly to a response by the mouse by changing the test concentration within a window of two. Thus, these single simulations suggest that the randomized window method brings animals to criterion faster than the posterior likelihood method, as hypothesized.

To characterize its performance more thoroughly, we completed a more detailed simulation analysis of the randomized window method. As with the posterior likelihood method, the randomized window method (when simulated with 500 mice) was more accurate when a high threshold was guessed (Figures 4b and 5a). We then proceeded to test convergence of the program when the guess was 0 and the actual α varied throughout the entire range of odorant concentrations.

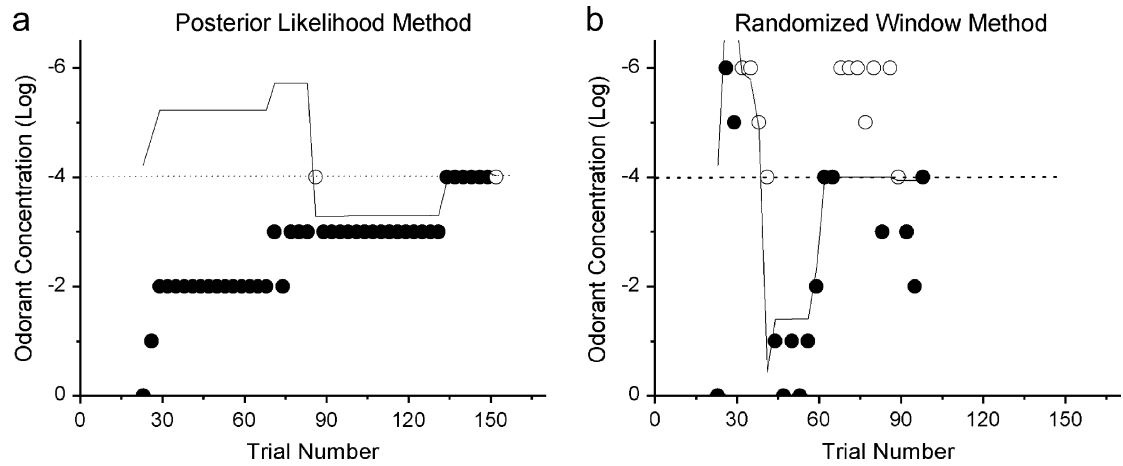


Figure 4 Simulation of single mouse when guessed threshold is 1% and actual threshold is $10^{-4}\%$ (shown as -4 , dotted line). Filled circles indicate positive responses ("licking"), and open circles indicate negative responses ("not licking"). Solid black line: α produced at the end of each trial. **(a)** Posterior likelihood method. **(b)** Randomized window method.

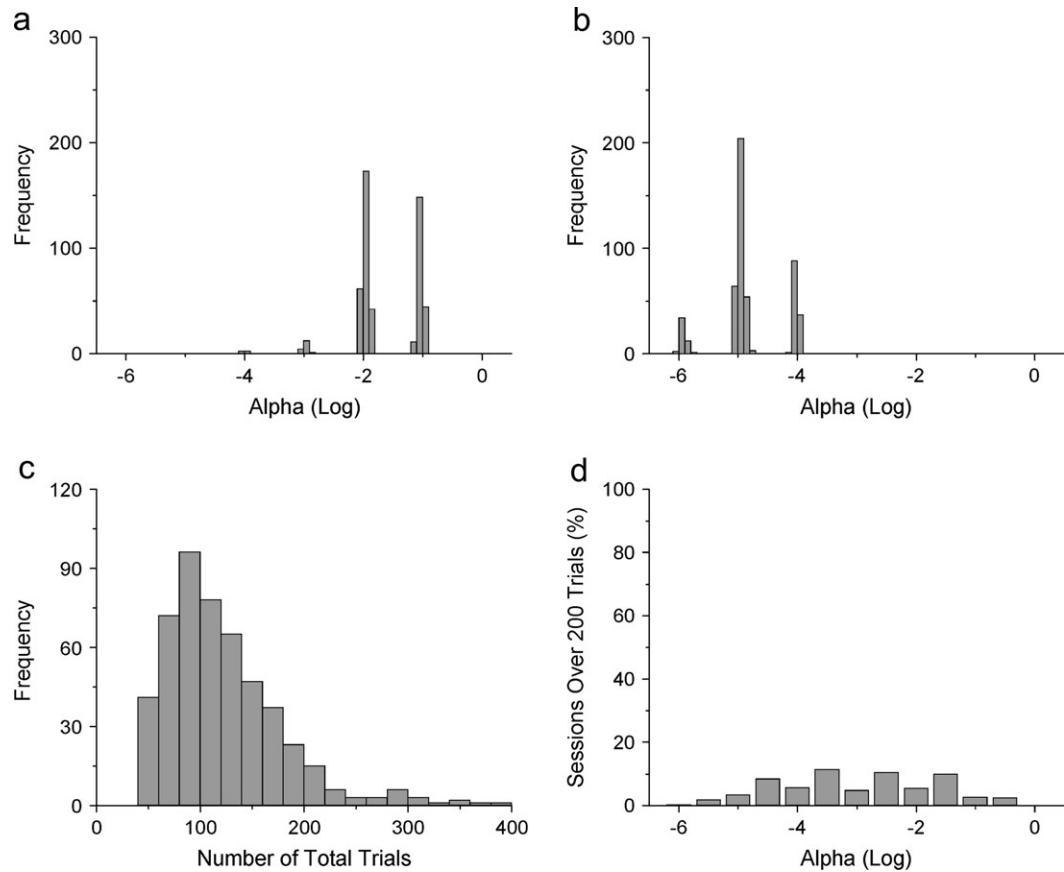


Figure 5 Randomized window method utilized in 500 simulation runs. **(a)** Histogram of α s (thresholds) produced when guessed threshold is $10^{-6}\%$ (shown as -6) and actual threshold is $10^{-1.5}\%$ (-1.5). The mean α was -1.63 , the median -1.9 , and the SD 0.55 . **(b)** Histogram of α s (thresholds) produced when guessed threshold is 1% (0) and actual threshold is $10^{-4.5}\%$ (-4.5). The mean α was -4.81 , the median -4.94 , and the SD 0.54 . **(c)** Histogram of the total number of trials to criterion for a guess α of 0 and actual α of -4.5 . The average number of total trials to criterion was 115 , and 6.4% of the 500 sessions required more than 200 trials to complete. **(d)** Percentage of runs that required over 200 trials to reach criterion when the guessed threshold is 1% and actual threshold varies from $10^{-0.5}\%$ to $10^{-6}\%$ (data are for a simulated observer with $\beta = 0.6$). When β for the simulated observer was changed to 3.5 , there was a significant increase in the number of runs over 200 . Thus, for example, for a true α of -2.5 , the number of runs over 200 trials was 92% .

We found that the randomized window method converged much faster when the difference between the guessed and actual thresholds was either small or large (Figure 5d). For example, with a guessed threshold of 0 and an actual threshold of -6 , the number of animals requiring more than 200 trials to reach criterion was only 0.2% compared to 81.2% for the posterior likelihood method.

We examined the bias and the variability for the estimated α for the two methods for choosing the next test stimulus. Bias was determined as the difference between the true and mean estimated α , and the variability was measured by the SD or the 68% CI for the estimated α . For a Gaussian distribution, the half width of the 68% CI is equal to the SD. We found that both the posterior likelihood and randomized window methods deviated systematically from the true α , but in opposite directions (Figure 6a). Thus, the posterior likelihood method yielded values of α that were higher than the true α . In contrast, the randomized window method resulted in an estimate of α that was smaller than the true α . In addition, when the true threshold was set to a half-log unit (i.e., at values of -0.5 , -1.5 , etc) the half width of the 68% CI was the same for both methods (Figure 6c), and there was only a small difference between the SDs (Figure 6b). At whole log units (i.e., -1 , -2 , etc), these measures of variability differed substantially between methods (Figure 6b,c). Because biological variability is such that thresholds rarely lie precisely at whole log units, the end result is that the variability of the posterior likelihood method is slightly lower than the variability of the randomized window method. In conclusion, the simulation experiments indicated that the randomized window method converges to an estimate of α faster than the posterior likelihood method, but at the expense of a larger variability for the estimate of α .

Changing β in the MLPEST program affects the estimate of α slightly

We examined the effect of changing the value of β in the MLPEST program for both the posterior likelihood and randomized window methods. We ran 500 simulations in which the silicon mouse was set to perform with a $\beta = 0.6$ while the MLPEST program used a β of either 3.5 or 0.6. For these simulations we used a guessed threshold of 1% and an actual threshold of -4 . The results from the posterior likelihood method and randomized window method were similar (except for the slightly larger number of trials required by the posterior likelihood method). Because others have previously demonstrated that the value of β has little effect on the computed α by MLPEST using the posterior likelihood method (Terutwein and Strasburger, 1999; Linschoten *et al.*, 2001), we will discuss these data in terms of the randomized window method. As expected, the simulated mouse yielded exactly the same psychometric curve regardless of the choice of β for MLPEST (Figure 7a,d). This indicates that the program was able to reproduce the Weibull curve that reflected the silicon mouse's actual performance (which was kept at 0.6). When the program used a β of 0.6, the estimated α s were less variable than if the program had used a β of 3.5 (Figure 7b,e). Unfortunately, when a β of 0.6 was used by MLPEST, 100% of the simulation runs required more than 200 trials to reach criterion (Figure 7c). This problem was worse with the posterior likelihood method, regardless of β (results not shown). This was much higher than the 5.6% of runs which exceeded 200 trials when β was set to 3.5 (Figure 7f). Because this would prohibit virtually all real mice from reaching criterion in a single session, β was kept at 3.5.

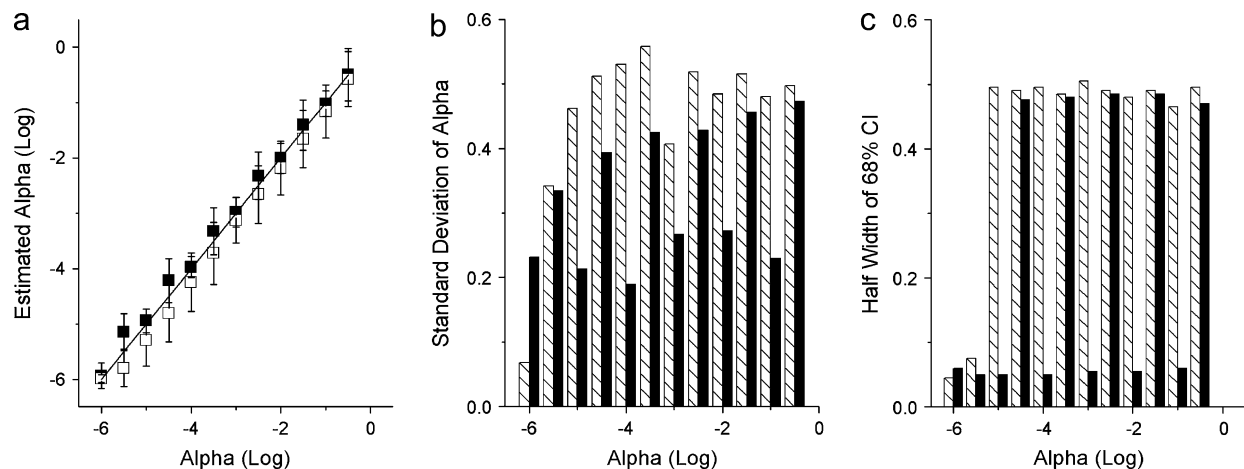


Figure 6 Bias and variability of the estimate of threshold (α). MLPEST simulations were run for 500 mice under each condition. The guessed threshold was 1%, and the actual threshold was varied from -0.5 to -6 in 0.5 log units. **(a)** Actual threshold versus average estimated threshold from 500 simulation runs. Solid black squares: posterior likelihood method; open squares: randomized window method; line: estimated threshold = actual threshold. Error bars indicate SD. **(b)** SD of the estimated threshold plotted as a function of actual threshold. Solid bars: posterior likelihood method; striped bars: randomized window method. **(c)** Half width of the 68% CI plotted as a function of actual threshold. Solid bars: posterior likelihood method; striped bars: randomized window method. The half width of the 68% CI would be equal to the SD if the variability of the threshold were Gaussian.

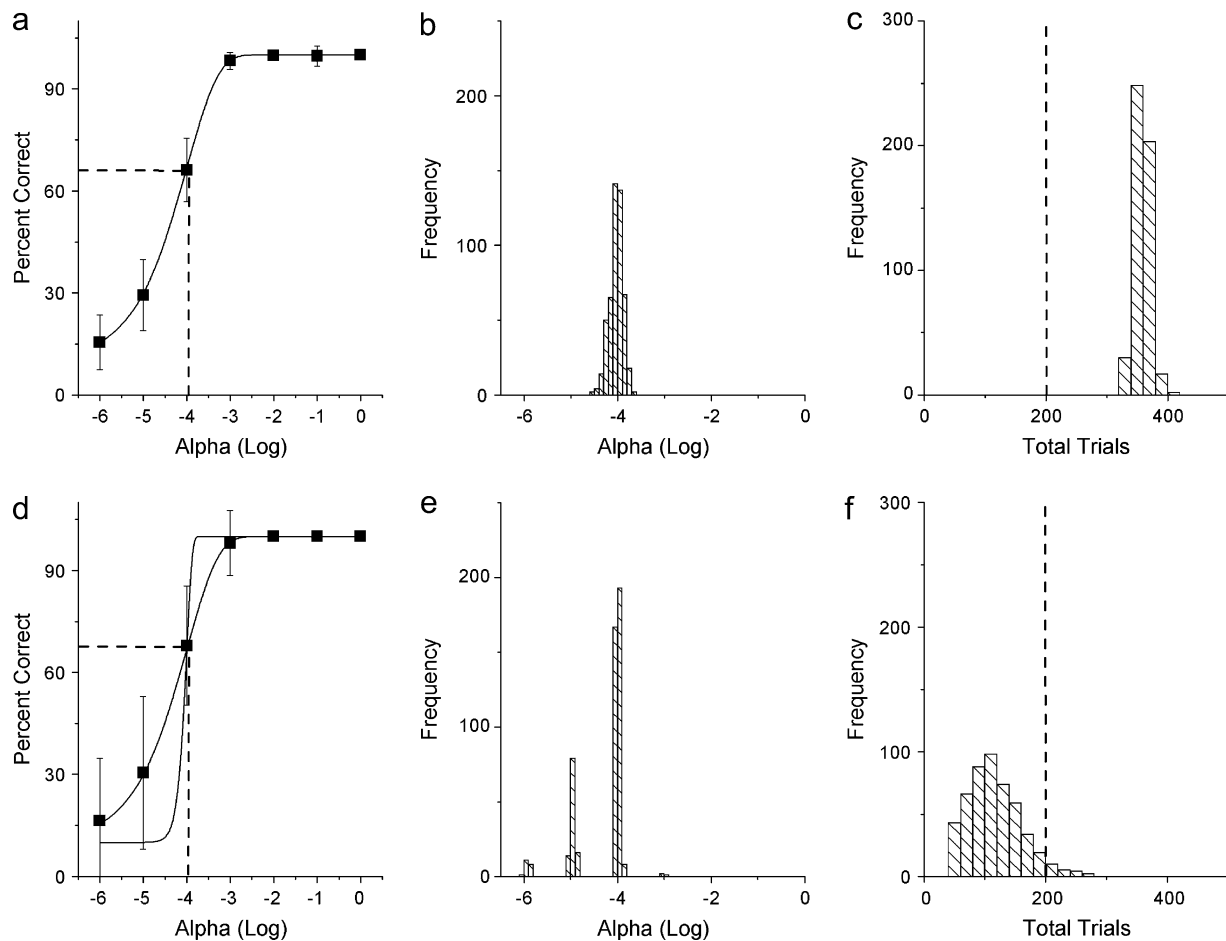


Figure 7 Simulation of randomized window method with different values of β in the MLPEST program. Guessed threshold was 0, and actual threshold was -4 . The silicon mouse always performs with a $\beta = 0.6$, while the β used in the MLPEST program is either 0.6 (**a, b, c**) or 3.5 (**d, e, f**). Five hundred simulated MLPEST sessions were run to generate the data shown in the figures. (a) Percent correct as a function of threshold (α , in log scale). MLPEST program run with $\beta = 0.6$. Squares are mean performance for 500 simulation sessions. Error bars indicate SD. Curve indicates Weibull function with a β of 0.6. Horizontal dotted line indicates 64% correct response. Vertical dotted line indicates actual threshold. (b) Histogram of α values computed when MLPEST is simulated using a β of 0.6 in the program and β of 0.6 for the simulated mouse. The mean estimated α was -4.03 , the median -4.01 , and the SD 0.15. (c) Histogram of the number of trials required to reach criterion when MLPEST is simulated using a β of 0.6. Hundred percentage of the sessions took over 200 total trials. (d) MLPEST program run with $\beta = 3.5$; simulated mouse has a β of 0.6. Squares are mean performance for 500 simulation runs. Error bars indicate SD. Solid black line is a Weibull function with a β of 0.6. Solid red line is a Weibull function with a β of 3.5. Horizontal dotted line indicates 64% correct response. Vertical dotted line indicates actual threshold. (e) Histogram of α values computed when MLPEST is simulated using a β of 3.5. The mean of the estimated α was -4.27 , the median -4.01 , and the SD 0.53. (f) Histogram of the number of trials required to reach criterion when MLPEST is simulated using a β of 3.5. A total of 5.6% of the sessions required over 200 total trials to reach the stopping CI.

MLPEST threshold estimates for CNGA2 KO animals

We used the MLPEST technique to obtain detection threshold estimates for CNGA2 KO mice and their WT littermates. Because we did not know what the odorant thresholds would be in these mice, we decided to use the randomized window method. Since the cAMP-mediated second messenger pathway was thought to be the primary signal transduction mechanism of most olfactory sensory neurons, we hypothesized that rendering mice deficient for this pathway would result in a higher detection threshold. Detection thresholds were determined using the MLPEST program (with the randomized window method) for EA, isoamyl acetate, octaldehyde, and 2-heptanone. 1% concentrations of each odorant were

used as the S+ control stimulus, and test concentrations ranged from $10^{-1}\%$ to $10^{-6}\%$. The S− control stimulus for all odorants was pure MO. Detection threshold did not differ between the KO animals and their WT controls for any of the four odorants tested (Figure 8). None of the thresholds differ between KO and control as indicated by $P > 0.10$ in a Wilcoxon test (P values are shown in the figure).

CNGA2 KO mice were also tested with geraniol. KO mice did not reach the 85% correct criterion in 200 trials when asked to detect 1% geraniol compared to MO, while WT mice reached criterion within six blocks of 20 trials (Figure 9). The reduced performance of KO animals in the detection test for geraniol made it impossible to conduct MLPEST for this

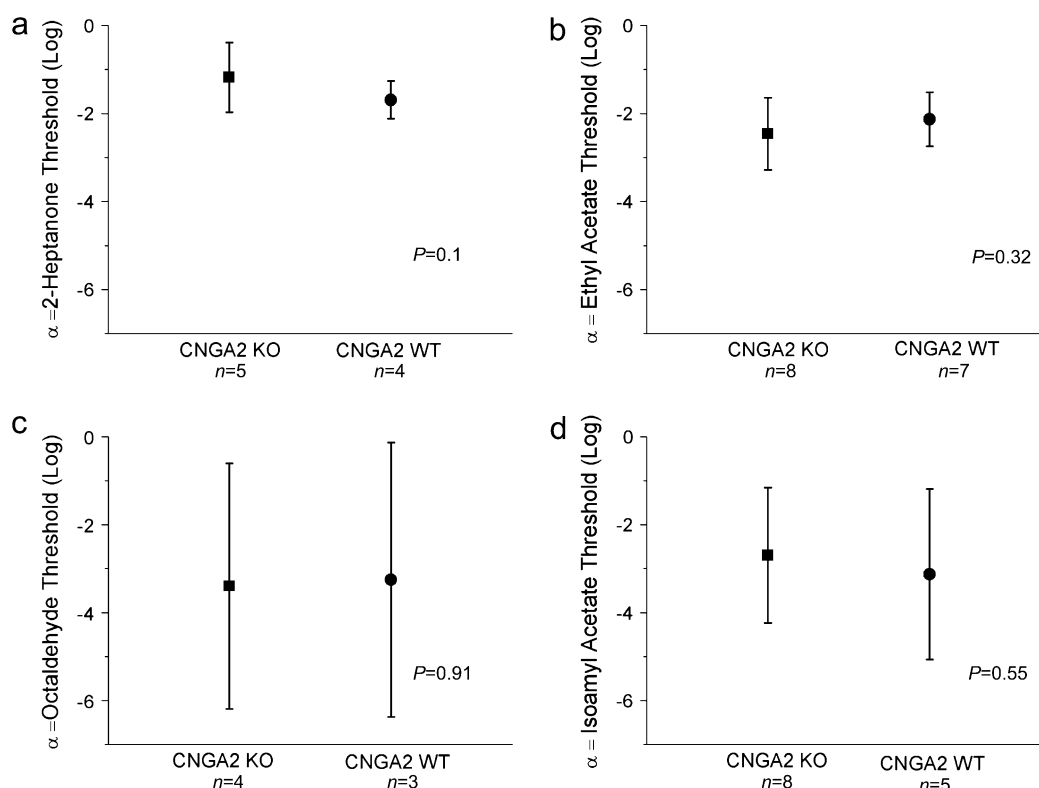


Figure 8 Mean detection thresholds and SDs for CNGA2 KO animals and WT littermates. Odorant concentrations ranged from 1% (shown as 0 on a logarithmic scale) to $10^{-6}\%$ (-6). Squares: CNGA2 KO animals; circles: CNGA2 WT animals. Errors bars indicate SD; n : number of animals that produced thresholds for each bar. P = P value for a test of the significance of the difference between the values for CNGA2 KO and controls using a Wilcoxon test. **(a)** 2-Heptanone KO animals: mean = -1.18 , median = -1 , and SD = 0.79 . WT animals: mean = -1.69 , median = -1.90 , and SD = 0.43 . **(b)** EA KO animals: mean = -2.46 , median = -2.94 , and SD = 0.81 . WT animals: mean = -2.13 , median = -1.95 , and SD = 0.62 . **(c)** Octaldehyde KO animals: mean = -3.40 , median = -4.55 , and SD = 2.79 . WT animals: mean = -3.25 , median = -5 , and SD = 3.12 . **(d)** Isoamyl acetate KO animals: mean = -2.70 , median = -2.95 , and SD = 1.54 . WT animals: mean = -3.12 , median = -4.24 , and SD = 1.94 .

odorant. MLPEST for geraniol was not run on either KO animals or WT controls.

Attempts to train animals at low odor concentrations

The estimate of threshold for EA obtained with MLPEST (-2.13) is three orders of magnitude higher than the estimate with the descending method of limits (-5) (compare Figures 2 and 8b). This called into question whether the higher threshold estimates with MLPEST were artifactual due, for example, to a more difficult MLPEST task compared to the regular S+/S− sessions run in the descending method of limits. If the threshold determined by MLPEST was higher than the true threshold, then newly trained mice that had achieved criterion at 1% EA on the second day of training should be successful in completing a go-no-go experiment at a concentration (-3) slightly below MLPEST threshold (-2.13), but higher than the threshold determined by the method of limits (-5).

Seven mice that had been successfully trained to discriminate 1% EA from MO were asked to complete the S+/S− paradigm using -3 EA as the S+ and MO as the S−. No mouse passed this paradigm on day 1. We repeated the experiment

for six more days. One animal passed the second day (on block 15); the remaining animals never passed, completing an average of 56 ± 2 blocks before testing was terminated. When we returned the animals to a higher concentration of EA, the animals needed 2 days of training to respond to the S+ odorant (-1 EA) and refrain from responding to the S− odorant (MO). These data indicate that the mice were not able to detect -3 EA versus MO unless substantial training had been completed using the schedule of odor concentrations used for the descending method of limits.

Maintenance of stimulus control

We found that the maintenance of stimulus control during MLPEST trials was fairly sensitive to the reward parameters. In the S+/S− program, mice are always rewarded on correct responses to S+ and never rewarded on S−. This reward paradigm was maintained in MLPEST. There was some discussion on how to reward the test concentrations, however. In this program, the false alarm rate is generally less than 20% and is often 0. We therefore considered it safe to reward test concentrations if the animal responded and initially maintained this reward schedule in MLPEST. When tested using

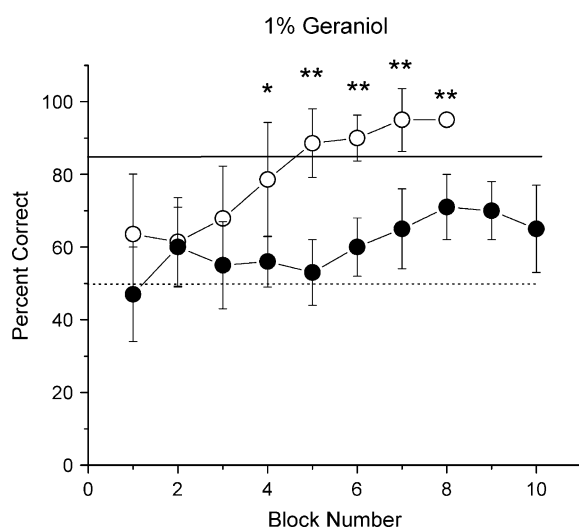


Figure 9 Detection of 1% geraniol in MO. Blocks indicate 20 S+/S− trials. To reach criterion, animals were required to maintain performance at or above 85% for three consecutive blocks. Data are given as a mean of the animal's performance for each block. Error bars indicate SD from the mean. Dotted line: chance performance; solid black line: criterion performance (85%); filled circles: CNGA2 KO animals ($n = 5$); and open circles: WT animals ($n = 7$). * $P < 0.05$ and ** $P < 0.01$ in a repeated measures ANOVA.

this reward schedule, however, animals very quickly lost stimulus control and began to respond to all trials regardless of whether odorant was present. In contrast, upon removal of the reward for test concentrations, mice stopped working—choosing instead to sleep or investigate the chamber. However, if test concentrations were rewarded randomly at a rate of 50%, mice maintained stimulus control and continued to work as long as no more than three test concentrations in a row were rewarded. This was therefore adopted as the final reward paradigm. In order to further strengthen stimulus control, mice were given four refresher S+/S− trials (without interspersed test trials) following any false alarm response during the MLPEST trials.

The combination of the water reward paradigm and the four refresher S+/S− trials following a false alarm maintained most animals under stimulus control. We discarded data for those few animals that lost stimulus control or whose psychometric function was not concentration dependent (which indicated that the animal cued to nonolfactory events such as valve sound when choosing to respond). For EA, 8 WT and 10 KO animals were tested; stimulus control and therefore results were obtained from 7 and 8 animals, respectively (1–2 animals per group were discarded). For 2-heptanone, six WT and nine KO animals were tested, with results obtained from four and five animals, respectively (two to four animals per group were discarded). For isoamyl acetate, five WT and nine KO were tested; results were obtained from five and eight animals, respectively (zero to one animal per group was discarded). For octaldehyde, five WT and seven KO animals were tested, with results obtained from three and four animals, respectively (two to three ani-

mals per group were discarded). Of the 15 total discarded data sets (out of 56 tested), 6 were discarded for a flat psychometric function and 9 were discarded due to imperfect stimulus control.

Psychometric function and false alarm rate

We ran several *post hoc* analyses to confirm that results by the CNGA2 KO mice were similar to the simulated MLPEST results. All analyses were completed for both CNGA2 KO mice and their WT littermates. There were no differences between KO and WT animals, so they will be discussed as a single group. We analyzed three facets of the results.

First, the final threshold given by the MLPEST program should not be affected by the number of test trials the animal went through (as indicated in Materials and Methods, each test trial was accompanied by one S+ trial and one S− trial given in random order). Accordingly, we found no correlation (Figure 10a), and the distribution of the data obtained in experiments with mice (Figure 10a) resembled the distribution obtained in a simulation experiment (Figure 10b).

Second, MLPEST methodology is based on the Weibull function. Although the MLPEST procedure is not designed to produce a full psychometric function, we thought it would be instructive to plot the average data for correct responses for each concentration for all animals. The combined percent correct responses as a function of odor concentration were analyzed for the closeness of fit to the Weibull curve. As expected, there was a large variation for correct response at each concentration (presumably not only due to biological variability but also due to the fact that MLPEST does not sample each concentration uniformly for each animal). When grouped by response to a particular odorant concentration, and after transformation of the data to make $\delta = 0$, the resulting curve was fit adequately by a Weibull curve, where β was allowed to vary and α was the mean of the value determined from the MLPEST sessions (−2.13 for WT mice and −2.46 for CNGA2 KO). Estimated β values fell between 0.7 and 0.5 (Figure 10c). Because there is significant variability in the value of α (Figure 10a), these population estimates for β are lower limit estimates of β for the psychometric function describing the behavior of individual mice.

Third, we examined the potential variation in false alarm rate throughout MLPEST trials. Compared to bias-free procedures such as the two-alternative forced choice (2AFC), the go-no-go paradigm utilized in the Bodyak and Slotnick olfactometer suffers from the potential problem that the false alarm rate could vary as a function of odorant concentration, thereby affecting the estimation of α (Harvey, 1986). This would be particularly problematic if the false alarm rate changed rapidly depending on the last concentration tested. If this were the case, false alarm would differ systematically for those tests following an S+ trial compared to those following an S− trial. We sorted the false alarms from all experiments depending on the previous stimulus to determine whether this was a problem. We found no

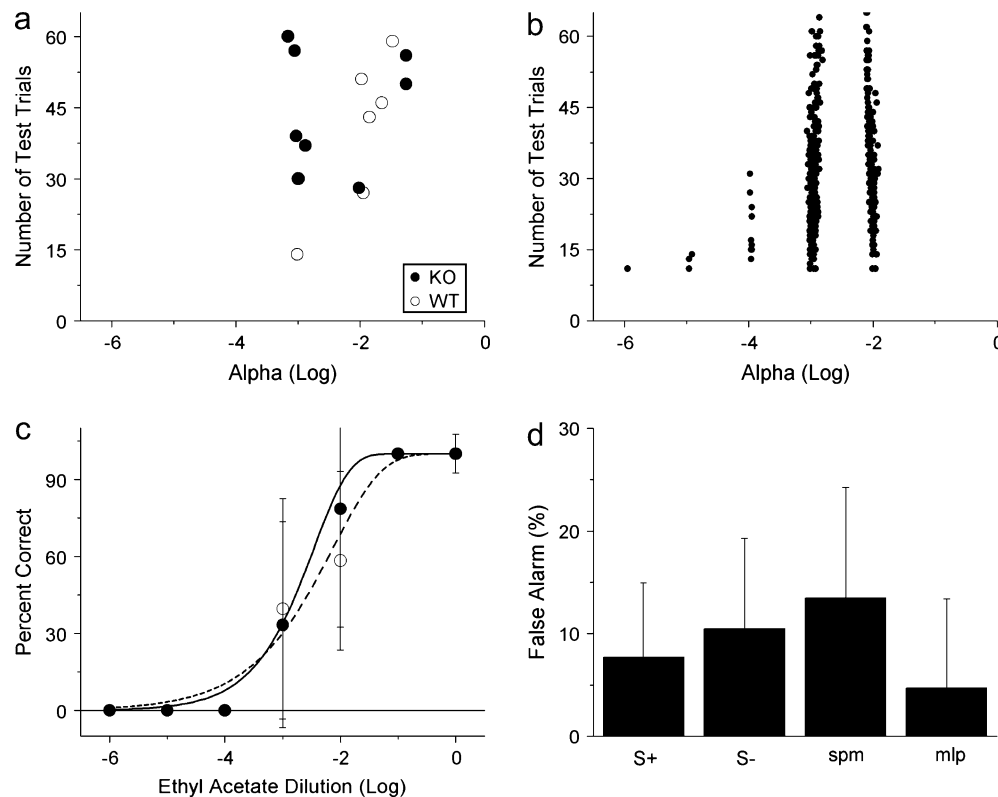


Figure 10 Post hoc analyses of EA MLPEST results for CNGA2 animals. Filled circles: CNGA2 KO animals; open circles: WT animals. **(a)** Number of MLPEST test trials to reach criterion versus α threshold for CNGA2 KO and control mice. The average number of test trials to completion for all mice combined was 41.8. **(b)** Number of MLPEST test trials to reach criterion versus α threshold for 500 simulated mice determined under the same starting conditions as in (a) (guessed $\alpha = 0$, with true $\alpha = -2.5$, $\beta = 0.6$ for the simulated mouse). The mean estimated α was -2.63 , and the average number of trials was 33. When the same simulation was run with a simulated mouse with $\beta = 3.5$, the estimated α was -2.51 and the average number of trials to completion was 93.3. **(c)** Mean response for each odorant concentration during MLPEST. Solid line: Weibull curve that fits CNGA2 KO animal data ($\beta = 0.7$); dashed line: Weibull curve that fits CNGA2 WT animal data ($\beta = 0.51$). Error bars indicate SD. The data in this plot have been transformed to make $\delta = 0$ using the equation $(PC - \delta)/(100 - \delta)$, where PC is percent correct. **(d)** False alarm rate. S+: false alarm rate following an S+ stimulus (1% EA in MO). S-: false alarm rate following an S- stimulus (MO alone). spm: false alarm rate during S+/S- (the first 20 trials). mlp: false alarm rate during MLPEST trials (all trials after the first 20).

dependence of false alarm on whether the previous trial was S+ or S- (Figure 10d, comparing bars for S+ and S-, there was no significant difference between any of the values shown). In addition, we tested for slower changes in false alarm rate by comparing the false alarm rates calculated during the first 20 S+/S- trials with false alarm rates measured during the rest of the trials in all MLPEST experiments. These were not significantly different either (Figure 10d, compare spm and mlp; an ANOVA yields no significant difference for any of the average values in this figure, $P > 0.09$).

Estimation of threshold with MLPEST in the presence of a background odor and stability between sessions

Figure 11 shows the estimation of threshold for heptaldehyde at concentrations ranging from -1.5 to -4.5 in 0.5 log steps in the presence of -1 octaldehyde. Each mouse was run through three MLPEST sessions. At a glance, the data for threshold as a function of session (Figure 11a) do not reveal any systematic change in threshold across sessions. Indeed, the correlation coefficient of 0.16 was not significantly different from zero

($P = 0.62$) and a Friedman's ANOVA did not find a difference of the threshold between sessions ($P = 0.85$). Similarly, there was no systematic change in the number of test trials per session (Figure 11b). The regression coefficient for the number of test trials as a function of session was 0.12 , not significantly different from zero with $P = 0.7$, and the ANOVA did not find statistical support for differences in the number of test trials per session between sessions $P = 0.37$. Thus, with the limited data we present, it appears that estimates for threshold are stable across sessions. Since this experiment utilized only four mice, further work is necessary to substantiate stability of the MLPEST threshold. The ability to determine a threshold utilizing odor mixtures as stimuli could be used in the future to determine discrimination thresholds with MLPEST.

Discussion

The accurate determination of olfactory thresholds is key to understanding the cellular and molecular mechanisms underlying olfaction. Unfortunately, likely due to the difficulty of performing the measurement, only a few studies have

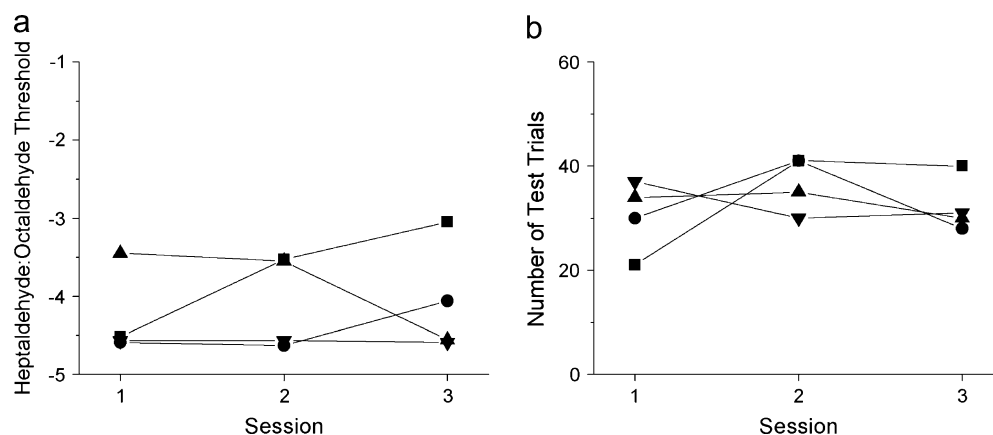


Figure 11 Results of MLPEST threshold determinations for heptaldehyde in three separate sessions in the presence of a background odor (octaldehyde). Four mice were run through three separate MLPEST sessions. The stimuli were [odor 1 (log concentration of odor 1): odor 2 (log concentration of odor 2)]: S+: octaldehyde (−1)/heptaldehyde (−1.5). S−: octaldehyde (−1). MLPEST stimuli: heptaldehyde:octaldehyde concentrations of: (−1.5):(−1), (−2):(−1), (−2.5):(−1), (−3):(−1), (−3.5):(−1), (−4):(−1), and (−4.5):(−1). **(a)** Threshold value (expressed as the logarithm of heptaldehyde concentration) as a function of session number. The data for each mouse are represented with different symbols and are linked with lines. The mean thresholds and SDs for each session were -4.28 ± 0.56 , -4.07 ± 0.61 , and -4.07 ± 0.72 . The median thresholds were -4.54 , -4.06 , and -4.31 . **(b)** Number of test trials required for convergence of the MLPEST program is shown for each mouse as a function of session number. The mean numbers of trials to criterion and SDs were 30.5 ± 7 , 37 ± 5 , and 32 ± 5 , and the medians were 32, 38, and 30.5.

attempted to determine odor thresholds in mice, often with a small number of animals per experiment (Youngentob and Margolis, 1999; Vedin *et al.*, 2004; Youngentob *et al.*, 2004; Pho *et al.*, 2005). These studies have used the descending method of limits to determine threshold. In this article we implement a promising new method (MLPEST) for use in determining odor detection thresholds in mice in the Bodyak and Slotnick olfactometer. We find that MLPEST provides a reliable and fast determination of threshold in mice in a single session. However, thresholds determined by MLPEST are two orders of magnitude higher than thresholds estimated by the descending method of limits, and we discuss the potential reasons for this discrepancy below (see “Why is the estimate of the threshold different between methods?”). Both MLPEST and the descending method of limits have significant potential limitations, and therefore, further future work is necessary to fully validate the methods for determination of detection and discrimination thresholds in mice. Utilizing the MLPEST program, we find that for the odors they can detect, CNGA2 KO mice have similar detection thresholds to WT controls.

MLPEST yields faster determination of threshold

The experiments completed demonstrate that MLPEST is a suitable technique for the rapid determination of olfactory detection threshold in mice. For all odorants tested, an average of 75% of the mice tested were able to complete MLPEST trials to threshold, in a single day, within the 0.5 CI and while remaining under stimulus control. Starting at the same concentration (1%), these animals would have required seven or more days (depending on the odorant and individual animal) to produce a threshold using the one-up-two-down variant of the staircase method and 2–6

days using the descending method of limits (Figure 2). Thus, even when taking into account the fact that only 75% of the mice complete MLPEST trials, this method provides faster threshold determination than other methods.

As explained in the Introduction, MLPEST is a method that attains fast and accurate convergence to a threshold by choosing the concentrations to be tested based on a maximum likelihood fit of the Weibull function to the previously acquired data (Harvey, 1986, 1997; Linschoten *et al.*, 2001). In addition, MLPEST also provides a stopping criterion by determining when a CI for the estimated threshold is smaller than a user-specified stopping value. It is important to note that our group is not the first to attempt to implement maximum likelihood procedures to determine detection thresholds quickly and accurately to estimate olfactory thresholds in rodents. Youngentob and co-workers (1997) successfully used in rats a modification of a maximum likelihood method termed QUEST (Watson and Pelli, 1983). The modification consisted of using the maximum likelihood stopping criterion provided by QUEST, while using a “tracking procedure” (not involving maximum likelihood methods) to determine the threshold as well as the next stimulus. The modification of QUEST formulated by Youngentob and co-workers was not validated using Monte Carlo methods, but use of the method yielded accurate one-session threshold determination in rats. Their tracking procedure requires finding a concentration range (which they term the “dynamic interval”) below which the animals respond with no correct response. Since mice often perform with nonzero false alarm rates, the dynamic interval cannot be determined in go-no-go experiments with mice in the Bodyak and Slotnick olfactometer. Because of the difficulty in implementing the tracking procedure with mice and the lack of Monte Carlo validation,

we chose not to use this method, but rather resorted to using MLPEST, an improved method based on QUEST, that has been thoroughly tested by Harvey's group and others through Monte Carlo simulations and experiments in the visual system as well as in human olfaction and taste. A description of the modifications performed to QUEST to formulate MLPEST is provided by Harvey (1986).

MLPEST provides an accurate estimate of threshold for an idealized observer

In Monte Carlo simulations, the accuracy of threshold determination is assessed by determining the bias (deviation of the mean estimated threshold from true threshold) and the variability of the estimated threshold. Our results show that MLPEST yields estimates of α that are slightly biased for both the posterior likelihood method and the random window method (Figure 6a). This result is not surprising since previous work has shown that MLPEST produces slightly biased estimates of α with the posterior likelihood method (Linschoten *et al.*, 2001). Importantly, the studies performed by Linschoten and co-workers (2001) show that MLPEST yields smaller bias than two other methods of threshold determination: the ascending method of limits and the staircase procedure. Interestingly, the posterior likelihood and random window methods are biased in opposite directions (Figure 6a). Thus, in the future it might be possible to render MLPEST estimates unbiased by combining the two methods.

In terms of the variability, MLPEST methodology was found by Linschoten and co-workers (2001) to produce smaller CIs than the staircase procedure and the ascending method of limits, giving it higher precision. In this article we tested two methods for choosing the next stimulus in MLPEST. Both methods yield similar CIs for threshold values that fell in between the tested odor concentrations (Figure 6c, points at -0.5 , -1.5 , etc). However, for thresholds equal to the odor concentrations used in the determination of threshold, the posterior likelihood method is more accurate (Figure 6c, points at -1 , -2 , etc). Because thresholds rarely fall exactly at one of the concentrations chosen for MLPEST, the end result is that the posterior likelihood method yields a slightly lower variability. The major difference between the two methods is the number of trials required to reach criterion, with the randomized window method requiring fewer trials, even at concentrations where both methods yield the same CI (e.g., -5.5 , see Figures 3d, 5d, and 6c). The increased speed of convergence takes place at a cost of slightly decreased accuracy for threshold determination for the randomized window method compared to the posterior likelihood method. In mouse experiments where the thresholds are unknown or are suspected to be far from the S+ positive control stimulus, the randomized window method should be used. Otherwise, the posterior likelihood method is a better choice as it provides improved accuracy for threshold determination.

Influence of other parameters of the Weibull curve on estimation of threshold

Maximum likelihood adaptive methods estimate the threshold (α), which is one of the four variables in the Weibull psychometric function, while the other parameters (steepness β , false alarm γ , and lapse δ) are typically held constant. Only one parameter is estimated in adaptive methods such as MLPEST because the method can become completely unstable at the beginning of MLPEST trials if even two parameters (e.g., threshold α and steepness β) are being estimated. In Figure 10, we obtained estimates between 0.5 and 0.7 for the steepness of the Weibull curve (β) for the psychometric function for the entire population of mice. Any biological variability in threshold values between mice will lead to a decrease in the steepness (β) for the population psychometric function. Thus, the values of β estimated from Figure 10 should not be taken as an estimate of the true β for individual mice, but rather as a "lower limit" for β . We found that using the value of 0.6 for β in the MLPEST program resulted in the accurate determination of threshold, but unfortunately at the expense of very slow convergence (Figure 7). Because of this, we used a value of 3.5 for β in our MLPEST estimation of α . This steeper slope resulted in a faster determination of threshold, as expected from the fact that for a steeper slope the likelihood function will be narrower as a function of candidate α values. In addition, we performed simulations that show that the choice of the particular value of β used in the determination of α does not change the estimated value of α . These data are consistent with the earlier work showing that there is virtually no bias introduced in the determination of thresholds by having MLPEST steepness different from the "true" value set in the Monte Carlo observer (Terutwein and Strasburger, 1999).

Differences between the values of the false alarm rate (γ) and the lapse rate (δ) used in the MLPEST program and the values for these two variables in the mouse's Weibull function could result in significant bias in the estimation of α (Harvey, 1986; Wichmann and Hill, 2001). We show that the false alarm rate does not change as a function of the concentration of odor in the previous trial (Figure 10d). Even if the false alarm rate were to change, our version of MLPEST would minimize the resulting error because it uses a running average of the false alarm rate for calculations. As a result, the program would track changes in false alarm rate and provide the desired correction. If the variability in the false alarm rate is problematic, a 2AFC paradigm should be used instead of the go-no-go paradigm. Finally, the difference between the percent correct response at infinitely high odor concentrations and 100% (δ), a measure of "lapse" by the observer, was set to zero for our experiments because the mice rarely missed responding at high concentrations. If mice were lapsing significantly, this would result in a biased estimate of α . In that case,

it would be useful to test whether adjusting δ on a trial-by-trial basis, in the same manner that γ was adjusted in this study, would result in an accurate determination of threshold.

Failure of threshold estimation in a fraction of the mice

The fact that a fraction of the animals does not reach criterion for threshold determination in MLPEST (25%) or the descending method of limits (40%) is problematic. Some of these mice respond to nonchemosensory cues and yield a flat psychometric function in MLPEST or respond to MO versus MO in the descending method of limits. We are currently implementing improvements in the olfactometer to minimize nonchemosensory cues. Auditory cues can be minimized by enclosing the valves in a sound-reducing enclosure such as a Plexiglas box. Somatosensory cues could be transmitted to the mouse through subtle differences in airflow through each valve and through vibration transmitted through the floor of the operant chamber. We find that cutting the vibrissae reduces responses to airflow in a subset of mice. Finally, physically separating the operant chamber from the rest of the assembly of the olfactometer might mitigate transmission of high-frequency vibration cues through the floor of the chamber. A second factor that contributes to failure of mice in the MLPEST task is loss of stimulus control. Use of a 2AFC test instead of the go-no-go paradigm might alleviate this problem. Finally, uniform partial reinforcement for S+ and test trials might produce improved performance.

Why is the estimate of the threshold different between methods?

It should be noted that comparison of threshold values determined using different methods is problematic. One problem is that most threshold determinations in rodents have been performed with different odorants and that in some cases air dilution olfactometry is used while in other cases liquid dilution is used (Pietras and Moulton, 1974; Slotnick and Schoonover, 1993; Youngentob *et al.*, 1997; Apfelbach *et al.*, 1998; Youngentob and Margolis, 1999; Vedin *et al.*, 2004; Pho *et al.*, 2005). Another problem is that each method defines the criterion for threshold differently. Thus, for the MLPEST method used in our experiments (using a Weibull psychometric function), threshold was defined as the concentration of odorant that yielded 64% correct response. Alternatively, if a Gaussian psychometric function (which is symmetric while the Weibull function is not) were used to predict threshold in MLPEST, the threshold would have been defined as the odorant concentration yielding 50% accuracy and thus would predict a lower threshold than the Weibull MLPEST.

Nevertheless, it is instructive to compare threshold values that have been obtained with different methods, but with the same odorant and the same olfactometer. Surprisingly, we show that MLPEST yields an estimate of detection threshold

for EA three orders of magnitude higher (-2.13) than the threshold measured by us in the same olfactometer using the descending method of limits (-6). An even lower threshold (-8) was found in mice by Vedin and co-workers (2004) using the descending method of limits in the Bodyak and Slotnick olfactometer. Why are these values so different? One explanation is that MLPEST misses estimation of the low threshold simply because of the difficulty of the task (i.e., because the mouse has not reached a suitable attentional status). It could be argued that if this were the case, a newly trained mouse that has attained criterion at 1% EA should be able to respond correctly to a concentration slightly below MLPEST threshold (10^{-30} % EA) in subsequent S+/S− go-no-go sessions. As reported in Results, we find mice cannot make this large leap and fail to attain criterion even after considerable training at 10^{-30} %. This suggests that the threshold determined by MLPEST is accurate, but it may be that training at the lower concentration (10^{-30} %) does not result in proper stimulus control. Therefore, further experimentation needs to be done to provide solid evidence for the accuracy of MLPEST threshold determination in mice.

An alternate explanation for the difference in threshold between the two methods is that these differences are real. Because mice spend several days detecting odor stimuli at increasingly lower concentrations while undergoing the descending method of limits, changes in neuronal plasticity could take place in the olfactory sensory pathway—rendering them more sensitive to the odor tested. Indeed, recent work in our laboratory comparing naive, water-deprived, untrained mice with mice that have undergone threshold estimation through the descending method of limits show that the two groups of mice (trained and naive) differ substantially in OB responsiveness to the same odor (odor responsiveness was measured through *c-fos* up-regulation in periglomerular cells; Salcedo *et al.*, 2005). While the processes that lead to plastic changes in the OB upon training with the descending method of limits are not understood, plasticity is likely to be caused by modification of circuitry through centrifugal modulation, likely to effect not only the OB but also higher order processing centers (Wilson and Stevenson, 2003; McLean and Harley, 2004). Thus, MLPEST and the descending method of limits could be reporting on thresholds for animals under a different sensory perceptual status. This kind of training-induced plastic changes in sensory perception has been found in the somatosensory, visual, and auditory systems where they are named “perceptual learning” (Lu and Dosher, 2004; Witte and Kipke, 2005). A better understanding of the causes of the differences between these two threshold methods requires future experiments.

CNGA2 KO detection thresholds

It is surprising that CNGA2 KO animals have similar detection thresholds compared to their WT littermates for those

odors they can detect. Although surprising, this is yet one more example of the resiliency of the olfactory system to tampering. Other examples include mice in which 95% of the OE has been lesioned with methyl bromide; lesioned mice do not differ from unlesioned littermates in detection threshold for propanol, D-limonene, and ethyl acetoacetate (Youngentob *et al.*, 1997). Similarly, removal of a portion of the OB that includes the main area responsive to propionic acid does not affect the animal's detection of that odorant (Slotnick *et al.*, 1997).

A question raised by our results is how animals are able to maintain detection threshold when only a few glomeruli remain responsive to these odors (CNGA2 KO mice have two glomeruli responsive to EA rather than hundreds in WT mice; Lin *et al.*, 2004). One possibility is that the few remaining active glomeruli are those innervated by the olfactory sensory neurons bearing the highest affinity receptors for the compounds tested. A second possibility is that neuronal plasticity in the input to the OB in CNGA2 KO mice is able to compensate for the loss of olfactory ability. However, these changes are unlikely to be drastic changes in olfactory receptor gene expression or olfactory sensory neuron targeting because previous experiments show only minor changes in axon targeting in CNGA2 KO mice (Lin *et al.*, 2000; Zheng *et al.*, 2000). A third possibility is that the dramatically reduced olfactory input in CNGA2 KO mice could cause changes in signal processing in higher order processing akin to plastic changes observed by us in the OB of mice trained for threshold determination using the descending method of limits (Salcedo *et al.*, 2005). For example, Youngentob *et al.* (1997) postulate that a loss of lateral inhibitory mechanisms due to the loss of functional olfactory sensory neurons could account for increased ability to detect odorants in a severely damaged olfactory system. Presumably, without these lateral inhibitory mechanisms, mitral cells would respond to odorants at concentrations at which they previously would have been inhibited, producing an output that is behaviorally detectable (Youngentob *et al.*, 1997). Future experiments will be necessary to determine the factors that contribute to the maintenance of olfactory thresholds in CNGA2 KO mice.

Acknowledgements

CNGA2 KO mice were kindly provided by Dr John Ngai and Randall Reed. We would like to thank Dr Lewis O. Harvey Jr for providing us the MLPEST routines and for stimulating discussions, Dr Burt Slotnick for help setting up and running the olfactometers and discussion, and Drs Alan Spector, Ernesto Salcedo, and Robin Michaels as well as the anonymous reviewers for thoughtful discussion. Julie Arellano and Audrey Hathaway assisted with olfactometer training. This work was supported by National Institutes of Health grants DC00566, DC04657, MH068582 (D.R.) and F30 DC 5740 (A.C.C.).

References

- Apfelbach, R., Weiler, E., Asselbergs, W.F., Polak, E.H. and Slotnick, B. (1998) *Selective and reversible reduction of odor sensitivity in the rat by concanavalin A*. *Physiol Behav.*, 65, 513–516.
- Apter, A.J., Gent, J.F. and Frank, M.E. (1999) *Fluctuating olfactory sensitivity and distorted odor perception in allergic rhinitis*. *Arch. Otolaryngol. Head Neck Surg.*, 125, 1005–1010.
- Baker, H., Cummings, D.M., Munger, S.D., Margolis, J.W., Franzen, L., Reed, R.R. and Margolis, F.L. (1999) *Targeted deletion of a cyclic nucleotide-gated channel subunit (OCNC1): biochemical and morphological consequences in adult mice*. *J. Neurosci.*, 19, 9313–9321.
- Bodyak, N. and Slotnick, B. (1999) *Performance of mice in an automated olfactometer: odor detection, discrimination and odor memory*. *Chem. Senses*, 24, 637–645.
- Brunet, L.J., Gold, G.H. and Ngai, J. (1996) *General anosmia caused by a targeted disruption of the mouse olfactory cyclic nucleotide-gated cation channel*. *Neuron*, 17, 681–693.
- Buck, L.B. (2000) *The molecular architecture of odor and pheromone sensing in mammals*. *Cell*, 100, 611–618.
- Cain, W.S., Gent, J.F., Catalanotto, F.A. and Goodspeed, R.B. (1983) *Clinical evaluation of olfaction*. *Am. J. Otolaryngol.*, 4, 252–256.
- Cometto-Muniz, J.E., Cain, W.S. and Abraham, M.H. (2003) *Quantification of chemical vapors in chemosensory research*. *Chem. Senses*, 28, 467–477.
- Harvey, L.O. Jr (1986) *Efficient estimation of sensory thresholds*. *Behav. Res. Meth. Instrum. Comput.*, 18, 623–632.
- Harvey, L.O. Jr (1997) *Efficient estimation of sensory thresholds with ML-PEST*. *Spat. Vision*, 11, 121–128.
- Krantz, D.H. (1969) *Threshold theories of signal detection*. *Psychol. Rev.*, 76, 308–324.
- Lin, D.M., Wang, F., Lowe, G., Gold, G.H., Axel, R., Ngai, J. and Brunet, L. (2000) *Formation of precise connections in the olfactory bulb occurs in the absence of odorant-evoked neuronal activity*. *Neuron*, 26, 69–80.
- Lin, W., Arellano, J., Slotnick, B. and Restrepo, D. (2004) *Odors detected by mice deficient in olfactory cyclic nucleotide-gated channel subunit A2 stimulate the main olfactory system*. *J. Neurosci.*, 24, 3703–3710.
- Linschoten, M.R., Harvey, L.O. Jr, Eller, P.M. and Jafek, B.W. (2001) *Fast and accurate measurement of taste and smell thresholds using a maximum-likelihood adaptive staircase procedure*. *Percept. Psychophys.*, 63, 1330–1347.
- Lu, Z.L. and Doshier, B.A. (2004) *Perceptual learning retunes the perceptual template in foveal orientation identification*. *J. Vision*, 4, 44–56.
- Maloney, L.T. (1990) *Confidence intervals for the parameters of psychometric functions*. *Percept. Psychophys.*, 47, 127–134.
- McLean, J.H. and Harley, C.W. (2004) *Olfactory learning in the rat pup: a model that may permit visualization of a mammalian memory trace*. *Neuroreport*, 15, 1691–1697.
- Pho, V., Butman, M.L. and Cherry, J.A. (2005) *Type 4 phosphodiesterase inhibition impairs detection of low odor concentrations in mice*. *Behav. Brain Res.*, 161, 245–253.
- Pietras, R.J. and Moulton, D.G. (1974) *Hormonal influences on odor detection in rats: changes associated with the estrous cycle, pseudopregnancy, ovariectomy, and administration of testosterone propionate*. *Physiol. Behav.*, 12, 475–491.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992) *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK.

- Salcedo, E., Zhang, C., Kronberg, E. and Restrepo, D.** (2005) *Analysis of training-induced changes in ethyl acetate odor maps using a new computational tool to map the glomerular layer of the olfactory bulb.* Chem. Senses, 30, 615–626.
- Slotnick, B.M., Bell, G.A., Panhuber, H. and Laing, D.G.** (1997) *Detection and discrimination of propionic acid after removal of its 2-DG identified major focus in the olfactory bulb: a psychophysical analysis.* Brain Res., 762, 89–96.
- Slotnick, B.M. and Schoonover, F.W.** (1993) *Olfactory sensitivity of rats with transection of the lateral olfactory tract.* Brain Res., 616, 132–137.
- Terutwein, B. and Strasburger, H.** (1999) *Fitting the psychometric function.* Percept. Psychophys., 61, 87–106.
- Vedin, V., Slotnick, B. and Berghard, A.** (2004) *Zonal ablation of the olfactory sensory neuroepithelium of the mouse: effects on odorant detection.* Eur. J. Neurosci., 20, 1858–1864.
- Watson, A.B. and Pelli, D.G.** (1983) *QUEST: a Bayesian adaptive psychometric method.* Percept. Psychophys., 33, 113–120.
- Wetherhill, G.B. and Levitt, H.** (1965) *Sequential estimation of points on a psychometric function.* Br. J. Math. Stat. Psychol., 18, 1–10.
- Wichmann, F.A. and Hill, N.J.** (2001) *The psychometric function: I. Fitting, sampling, and goodness of fit.* Percept. Psychophys., 63, 1293–1313.
- Wilson, D.A. and Stevenson, R.J.** (2003) *Olfactory perceptual learning: the critical role of memory in odor discrimination.* Neurosci. Biobehav. Rev., 27, 307–328.
- Witte, R.S. and Kipke, D.R.** (2005) *Enhanced contrast sensitivity in auditory cortex as cats learn to discriminate sound frequencies.* Brain Res. Cogn. Brain Res., 23, 171–184.
- Xu, F.Q., Greer, C.A. and Shepherd, G.M.** (2000) *Odor maps in the olfactory bulb.* J. Comp. Neurol., 422, 489–495.
- Youngentob, S.L. and Margolis, F.L.** (1999) *OMP gene deletion causes an elevation in behavioral threshold sensitivity.* Neuroreport, 10, 15–19.
- Youngentob, S.L., Pyrski, M.M. and Margolis, F.L.** (2004) *Adenoviral vector-mediated rescue of the OMP-null behavioral phenotype: enhancement of odorant threshold sensitivity.* Behav. Neurosci., 118, 636–642.
- Youngentob, S.L., Schwob, J.E., Sheehe, P.R. and Youngentob, L.M.** (1997) *Odorant threshold following methyl bromide-induced lesions of the olfactory epithelium.* Physiol. Behav., 62, 1241–1252.
- Zheng, C., Feinstein, P., Bozza, T., Rodriguez, I. and Mombaerts, P.** (2000) *Peripheral olfactory projections are differentially affected in mice deficient in a cyclic nucleotide-gated channel subunit.* Neuron, 26, 81–91.

Accepted October 24, 2005