The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less)

Nicolaas Prins

Department of Psychology, University of Mississippi, Oxford, MS, USA

Adaptive testing methods serve to maximize the information gained regarding the values of the parameters of a psychometric function (PF). Such methods typically target only one or two ("threshold" and "slope") of the PF's four parameters while assuming fixed values for the "nuisance" parameters ("guess rate" and "lapse rate"). Here I propose the "psi-marginal" adaptive method, which can target nuisance parameters but only when this is the most optimal manner in which to reduce uncertainty in the value of the parameters of primary interest. The method is based on Kontsevich and Tyler's (1999) psi-method. However, in the proposed method a posterior distribution defined across all parameters of unknown value is maintained. Each of the parameters is specified either as a parameter of primary interest whose estimation should be optimized or as a nuisance parameter whose estimation should be subservient to the estimation of the parameters of primary interest. Critically, selection of stimulus intensities is based on the expected information gain in the *marginal posterior distribution defined across the parameters of interest only*. The appeal of this method is that it will target nuisance parameters adaptively and only when doing so maximizes the expected information gain regarding the values of the parameters of interest. Simulations indicate that treating the lapse rate as a nuisance parameter in the psi-marginal method results in smaller bias and higher precision in threshold and slope estimates compared to the original psi method. The method is highly flexible and various other uses are discussed.

Introduction

The psychometric function relates the performance of an observer to some quantitative stimulus characteristic (e.g., Kingdom & Prins, 2010). I will refer to the latter in the remainder simply as the *intensity* of the stimulus, symbolized by x. In models of observer performance, a specific shape of the function is typically assumed and is almost invariably some sigmoidal function such as the cumulative Gaussian distribution or the Weibull function. In the remainder, I will use $F(x; \alpha, \beta)$, or simply F, to refer to the function that characterizes performance of an underlying sensory mechanism. Two parameters characterize the function, its threshold and its slope. The threshold parameter (symbolized by *alpha*: α) specifies the function's location and is typically defined as that stimulus intensity at which a specific level of performance is achieved. The slope parameter (symbolized by *beta*: β) determines the function's rate of change. In this paper I use the Gumbel (or log-Weibull) function throughout as the model of F and use the subscript "G" to specify this:

$$F_G(x; \alpha, \beta) = 1 - \exp\left(-10^{\beta(x-\alpha)}\right) \tag{1a}$$

In practice, we cannot measure F directly. In psychophysical studies, we instead infer F from the pattern of responses (e.g., proportion correct in a multiple alternative forced-choice [mAFC] task or proportion endorsement [e.g., "I perceived the stimulus"] in a "yes/no" task) observed across trials in which stimuli of varying intensity x are used. The probability of observing a positive response as a function of stimulus intensity x is usually modeled as:

$$\psi(x; \alpha, \beta, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda)F_G(x; \alpha, \beta). \tag{1b}$$

F_G, x, α, and β are as defined above. The parameter γ corresponds to the lower asymptote of ψ, while the parameter λ determines the upper asymptote. Neither γ nor λ characterizes the performance of the underlying sensory mechanism, which is our primary interest. In a "yes/no" task, the lower asymptote, γ, corresponds to the false alarm rate and characterizes the decision process. In the more common mAFC task, γ is determined by the task and is assumed to equal $1/m$ in a task with m response categories. Parameter λ determines the upper asymptote of ψ. The upper asymptote often deviates from unity due to so called lapses, which are negative responses ("no" or incorrect) that are stimulus independent (for example, "finger errors").

Citation: Prins, N. (2013). The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less). *Journal of Vision*, *13*(7):3, 1–17, http://www.journalofvision.org/content/13/7/3, doi:10.1167/13.7.3.

doi: 10.1167/13.7.3 Received March 15, 2013; published June 7, 2013 ISSN 1534-7362 © 2013 ARVO

Thus, parameter λ also does not characterize the underlying sensory mechanism but rather is a function of observer characteristics such as attention or vigilance. Since γ and λ do not describe the underlying sensory mechanism but nevertheless do affect our observations (which we must model to recover the underlying function F), they are considered to be "nuisance parameters".

Much research has concerned itself with the optimal placement of stimuli: At which intensities x do we place stimuli in order to derive parameter estimates in a manner that is optimally efficient (i.e., requiring as few trials as possible in order to obtain parameter estimates of desired precision)? The answer to this question will depend on which of the PF's four parameters we are interested in. In order to estimate the stimulus intensity at which performance reaches some specific level, stimuli are most optimally placed at the intensity corresponding to the targeted performance level (e.g., Harvey, 1986; Levitt, 1971). In order to optimize slope estimation, stimuli should be placed at intensities that are some distance away from threshold intensity on either side (e.g., King-Smith & Rose, 1997; Levitt, 1971). If one is interested in estimating both threshold and slope, Levitt (1971) suggests using a compromise between the optimal placement for threshold estimation and optimal placement for slope estimation.

A complicating factor is that the nuisance parameters are correlated with both the threshold and slope parameters. Thus, an inaccurate model of the nuisance parameters (whether an incorrect assumed value or a poor estimate) will be associated with systematic misestimation of threshold and slope parameters. Use of an mAFC task gets around the issue as far as the guess rate is concerned. The lapse rate is not so easily dealt with and has been of concern in much research (e.g., Hall, 1981; Harvey, 1986; Klein, 2001; Manny & Klein, 1985; Prins, 2012a; Swanson & Birch, 1992; Taylor, Forbes & Creelman, 1983; Treutwein & Strasburger, 1999; Wichmann & Hill, 2001). It is often recommended to include a number of stimulus placements at a high intensity in case a free lapse rate parameter is included in the fitting procedure in order to improve the estimate of the lapse rate and thereby also improve the estimation of threshold and/or slope (e.g., Prins, 2012a, b; Swanson & Birch, 1992; Treutwein, 1995; Wichmann & Hill, 2001).

Adaptive placement methods utilize the information regarding the PF gained on preceding trials in order to optimize the placement of stimuli on subsequent trials. Adaptive methods can roughly be divided into two categories: parametric and nonparametric methods. Parametric methods (e.g., Kontsevich & Tyler, 1999; Pentland, 1980; Watson & Pelli, 1983; Watt & Andrews, 1981) assume a particular shape of PF and attempt to optimize estimation of one or more of its parameters. Nonparametric methods (e.g., Dixon & Mood, 1948; García-Pérez, 1998; Kaernbach, 1991; Taylor & Creelman, 1967; Wetherill & Levitt, 1965), on the other hand, assume no specific form of PF and instead optimize and target the estimation of one or more stimulus intensities that correspond to specific proportions correct/endorsement (i.e., values of ψ in Equation 1).

To date, no adaptive method addresses nuisance parameters in an adaptive manner. Nonparametric methods are concerned with estimating the intensity level at which a particular proportion endorsement (correct response or "yes") is obtained. Proportion endorsement confounds the performance of the underlying sensory mechanism with nuisance factors, such as lapses. Parametric adaptive methods generally assume a fixed, small value of the lapse rate and use a fixed, known value of the guess rate. As such, these methods do not place stimuli at intensities at which much information is gained regarding the lapse rate. Thus, even if observations are fitted in a separate procedure while allowing the lapse rate to vary, a poor estimate of the lapse rate is obtained and estimation of the parameters of interest is compromised (e.g., Prins, 2012a; Swanson & Birch, 1992). Inclusion of a number of stimuli at a very high intensity among trials otherwise controlled by an adaptive method improves estimates of the parameters of interest. Estimates are especially improved when "free trials" are placed at such intensities that it can safely be assumed that errors there can only be due to lapses, especially when the model that is fitted reflects this (e.g., Prins, 2012a; Swanson & Birch, 1992).

No clear guidelines exist as to what proportion of trials should be free trials in order to optimize the gain of information regarding the parameters of primary interest. Moreover, the optimal proportion of free trials will depend on many factors: e.g., which parameter(s) are of primary interest (threshold, slope, or both), the value of the actual, generating lapse rate, and the value and precision of estimates of *any* of the PF's parameters based on the information gained on previous trials. Including a fixed proportion of free trials among trials otherwise controlled by an adaptive method of course adds a distinct nonadaptive component into the stimulus selection procedure. Note that not only is the proportion of free trials not adjusted based on information gained on previous trials, the adaptive method in control of placement of the trials that are not free will also not modify its behavior based on information gained from the free trials.

Here I propose a modification of the adaptive psi-method (Kontsevich & Tyler, 1999) that will allow the procedure to target the nuisance parameters adap-

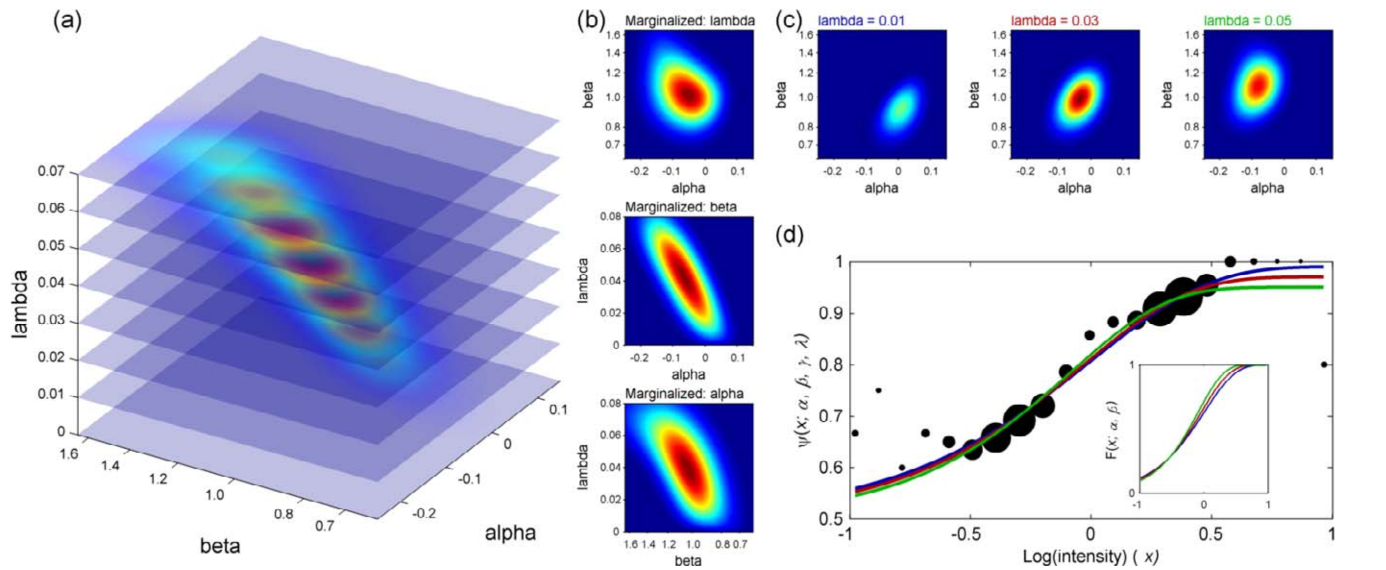

Figure 1. Dependencies among parameters using typical psi-method placement. (a) Posterior distribution across values of threshold (alpha), slope (beta), and lapse rate (lambda) based on the hypothetical data shown in (d). (b) The posterior distribution marginalized across lapse rate (top), slope (middle), and threshold (bottom). (c) Three constant lapse-rate slices ($\lambda = 0.01$, $\lambda = 0.03$, and $\lambda = 0.05$) through the posterior distribution. (d) Hypothetical data set ($N = 960$) in which stimulus placement corresponded to typical placement in psi-method runs and the pattern of responses is that which is most likely to be obtained with this placement regimen and generating function ($\alpha = 0$, $\beta = 1$, $\gamma = 0.5$, $\lambda = 0.03$) except for the occurrence of one incorrect response at the highest stimulus intensity used. Also shown are the uniform-prior Bayesian fits that assume $\lambda = 0.01$ (blue), $\lambda = 0.03$ (red), and $\lambda = 0.05$ (green). Note that while the modeled proportions correct (ψ in Equation 1) for all three correspond closely to each other and the data, the modeled values of the threshold parameter α, the slope parameter β, and function F (Equation 1) do not (Figure inset shows the three fitted functions in terms of $F[x;\ \alpha,\ \beta]$).

tively but only in order to maximize the efficiency of estimation of the parameters of primary interest. The psi-method, in its original form, targets threshold and slope estimates. It assumes a specific form for F (e.g., Weibull, Logistic, etc.), a fixed small value for the lapse rate (such as 0.03) as well as a value for the guess rate dictated by the experimental method (e.g., for a 2-AFC task a guess rate equal to 0.5 would be assumed). After each trial, the psi method updates a Bayesian posterior distribution across a range of possible values for the threshold and slope parameters of the PF. Based on this posterior distribution, the method determines for each of a limited number of possible stimulus intensities the probabilities of observing a positive and negative response. For each possible stimulus intensity, it also determines the entropy in the posterior distribution that would result for either of the two possible responses. The stimulus intensity chosen on the next trial is that intensity which minimizes the expected entropy in the posterior distribution. In other words, on each trial that stimulus intensity is selected, which is associated with the greatest expected information gain in the threshold and slope estimate.

When the assumed lapse rate matches the generating lapse rate, the psi-method recovers the threshold and slope parameters without bias and in a highly efficient manner (Kontsevich & Tyler, 1999). However, when assumed and generating lapse rate do not match, we can expect a systematic bias in both threshold and slope. The problem arises because, unless placement includes high intensities, the lapse rate shows a significant degree of dependency with threshold and slope parameters (Prins, 2012a). This is shown in Figure 1a, which displays the posterior distribution across 3-D parameter space (threshold $\times$ slope $\times$ lapse rate; guess rate is fixed at 0.5) for the hypothetical data shown in Figure 1d. Since a uniform prior was used across the parameter space shown, the density in the posterior is also proportional to the likelihood function. The placement of the 960 trials in this dataset was made to correspond to the average placement obtained across 2,000 simulated psi-method runs of 960 trials where the generating function was the Gumbel with threshold, slope, guess, and lapse rate values of 0, 1, 0.5, and 0.03 respectively. The lapse rate value assumed by the psi-method corresponded to the generating value. The areas of the symbols in the figure are proportional to the number of trials placed at the corresponding value of x. The proportions correct shown in Figure 1d correspond to the most likely to be obtained for this placement pattern and generating PF, except that one of the five trials presented at the highest intensity was set to be incorrect here (with this generating function it is in fact more likely for all five responses to be correct at this intensity rather than four).

The dependency among parameters is quite clear from Figure 1a in that the region of high density is slanted in parameter space. The dependency is clear also in Figure 1b, which displays the posterior distribution marginalized across each of the three parameters (threshold, slope, and lapse rate) individually. Figure 1c shows three slices through the posterior distribution corresponding to different values of the lapse rate ($\lambda = 0.01$, $\lambda = 0.03$, and $\lambda = 0.05$). Clear from these figures is that at high lapse rates, the posterior distribution peaks at lower thresholds and higher slopes than it does at low lapse rates. Using a fixed value of the lapse rate (as the psi-method does) corresponds to selecting one horizontal slice of the parameter space shown in Figure 1a. Our particular choice of value for the lapse rate will systematically affect our threshold and slope estimates. In other words, selecting the wrong value for the fixed lapse rate in the psi-method would lead to systematic bias in the threshold and slope estimates. Figure 1d shows the best-fitting PFs that assume the fixed values for the lapse rates shown in Figure 1c. While these functions are virtually identical with regard to modeled values of ψ at the intensities x at which observations are concentrated, the models differ with regard to the values of the parameters of interest (i.e., α and β). Modeled values of ψ for these three functions differ substantially only at high stimulus intensities, higher than those targeted by the psi-method.

In the following, I will first demonstrate that the psi-method results in biased estimates of the threshold and slope parameters when the assumed lapse rate and the generating lapse rate do not correspond. I will also demonstrate that this bias is not remedied by allowing the lapse rate to vary in a subsequent fitting procedure. Finally, I will propose a modification of the psi-method, the "psi-marginal method", that allows it to target nuisance parameters if doing so is the most efficient manner in which to reduce uncertainty in the parameters of primary interest. This is a general strategy which, depending on the researcher's goals and assumptions, allows any of the PF's four parameters to be treated either as a parameter of primary interest whose estimation should be optimized, a nuisance parameter whose estimation is subservient to the estimation of the parameters of primary interest or a fixed parameter of known value. Various examples will be presented. In all of the following, I will consider both Bayesian and maximum-likelihood (ML) parameter estimates.

Experiment 1: Bias in psi-method estimates when assumed lapse rate does not match generating lapse rate

Methods

Simulations

The generating function (F_G in Equation 1) in all simulations was a Gumbel ("log-Weibull") function with threshold $\alpha = 0$ and slope $\beta = 1$. Note that these specific values were chosen purely for mathematical convenience but that this choice of particular values has no consequence whatsoever for the conclusions drawn in this paper. The guess rate γ was set to 0.5 (effectively simulating a 2-AFC task; results for other values of γ are presented in the Supplementary materials). Three generating lapse rates were used: $\lambda = 0$, $\lambda = 0.025$, and $\lambda = 0.05$. Prior distributions in all simulations were uniform across a constrained 2-D (threshold $\times$ [log] slope) parameter space. The choice of limits on the parameter space will have a significant effect on Bayesian parameter estimates at the start of a run for as long as the posterior distribution would otherwise have mass beyond the edges of the prior distribution. However, as the run progresses the posterior distribution will become narrower and eventually will, effectively, be fully contained within the limits of the prior distribution. In order to control for the effect of the choice of limits of the prior, four different choices for limits on the prior across values for α and $\log(\beta)$ were used. These are illustrated in Figure 2. Each contained 21 threshold $\times$ 21 slope values. The threshold limits for the red and gray priors in the figure were $[F_{0.1}^{-1}, F_{0.9999}^{-1}] - 0.15$, those for the green and blue priors were $[F_{0.1}^{-1}, F_{0.9999}^{-1}] + 0.15$ (where F_p^{-1} is that stimulus intensity x at which the generating function $F_G(x; \alpha = 0, \beta = 1)$ evaluates to p). The (log) slope limits for the gray and blue priors were $[\log_{10}(0.0625), \log_{10}(16)] - 0.15$, those for the red and green priors were $[\log_{10}(0.0625), \log_{10}(16)] + 0.15$. For each of the three generating lapse rate values considered, 500 simulations were performed within each of the four groups of priors. For each of the 2,000 resulting simulations per generating lapse rate value, some random jitter (sampled from a rectangular distribution on [-0.15, 0.15]) was applied to the limits of the prior distribution. While mean parameter estimates will systematically differ between the four priors at a low number of trials, eventually the parameter estimates will converge when posterior distributions are (effectively) contained within the limits set by the prior. From that point forward, parameter estimates based on the constrained uniform priors will be, for all intents and purposes, equal to those based on priors which are truly uniform. The range of possible stimulus intensities the

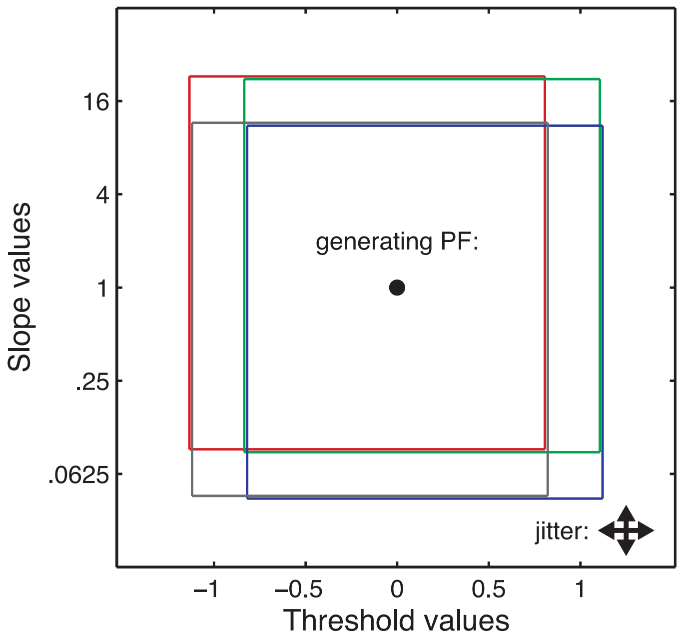

Figure 2. Schematic representation of the parameter ranges used in the priors. See text for detail.

psi-method could select from included 21 equal-spaced values between $F_{0.1}^{-1}$ and $F_{0.9999}^{-1}$. A total of 1,920 trials were simulated in each psi-method run. The psi-method was implemented using the Palamedes toolbox (Prins & Kingdom, 2009).

Parameter estimation

Results are analyzed and displayed in a manner similar to that used in Kontsevich and Tyler (1999). Bayesian threshold and slope estimates were derived after each trial in each run as the marginal means of the posterior distribution. Note that since the Gumbel function is used, stimulus intensities and threshold values are expressed in log units and an arithmetic mean on these values was computed. Slope values, however, are logarithmically spaced in the posterior and, correspondingly, a geometric mean was computed (as was done in Kontsevich & Tyler, 1999). Maximum likelihood (ML) estimates based on $N = 120$, 240, 480, 960, and 1,920 trials were also derived. The ML estimates correspond to those parameter values that maximize the likelihood function and were determined by an iterative simplex search (Nelder & Mead, 1965). Like the Bayesian estimates, ML estimates were derived using a fixed value for the lapse rate (0.03). Note that the maximum likelihood estimates were not constrained to lie within the limits placed on the prior. The standard error of threshold estimates for both Bayesian and ML estimates was derived as:

$$SE_{a_j} = \sqrt{\frac{\sum_I (a_{i,j} - \alpha)^2}{I}}, \tag{2}$$

where α is the generating value of the threshold (i.e., $\alpha = 0$), $a_{i,j}$ is the threshold estimate after Trial j in simulated run i, and I is the number of simulated runs (i.e., $I = 2{,}000$). The standard error of the slope was derived in log units:

$$SE_{b_j} = \sqrt{\frac{\sum_I \left(\log(b_{i,j}) - \log(\beta)\right)^2}{I}} \tag{3}$$

where β is the generating value of the slope (i.e., $\beta = 1$), $b_{i,j}$ is the slope estimate after Trial j in simulation i, and I is as above.

Results

Figure 3 shows the results for the simulations described above which, like the original psi-method, assumed a fixed lapse rate. The continuous lines show mean bias (left) and standard errors (right) in the Bayesian estimates for the three values of the generating lapse rate used across all 2,000 simulations (thick lines) as well as separately for the four different priors used (thin lines). Results from the four different priors that were used converge, roughly speaking, near Trial 100 or earlier. As argued above, this convergence signals the point at which the exact choice of limits on the prior no longer affects Bayesian parameter estimates (i.e., parameter estimates would no longer differ from those obtained with a nonconstrained uniform prior). Note that since the generating value for both α and $\log_{10}(\beta)$ were equal to 0, the mean bias is simply the mean parameter estimate itself. When the generating and assumed lapse rate coincide, I replicate Kontsevich and Tyler's (1999) result that, at least within a reasonable number of trials, the psi method is virtually free from bias. However, when the generating lapse rate deviates from that assumed by the psi-method, a systematic bias is obtained. Not surprisingly, the observed bias does not appear to resolve itself asymptotically. The degree of bias is relatively severe when compared against the standard error of estimate. The symbols in Figure 3 show obtained biases and standard errors in ML estimates. At high numbers of trials, Bayesian and ML estimates converge, as do their SEs. The counterintuitive initial rise in SE for slope was noted also by Kontsevich and Tyler (1999). At the start of a run, the posterior is flat with respect to slope and the slope estimate will simply be the center of the range of slope values contained within the prior. The initial rise in SE of slope occurs because at the start of a run the psi-method concentrates on the threshold rather

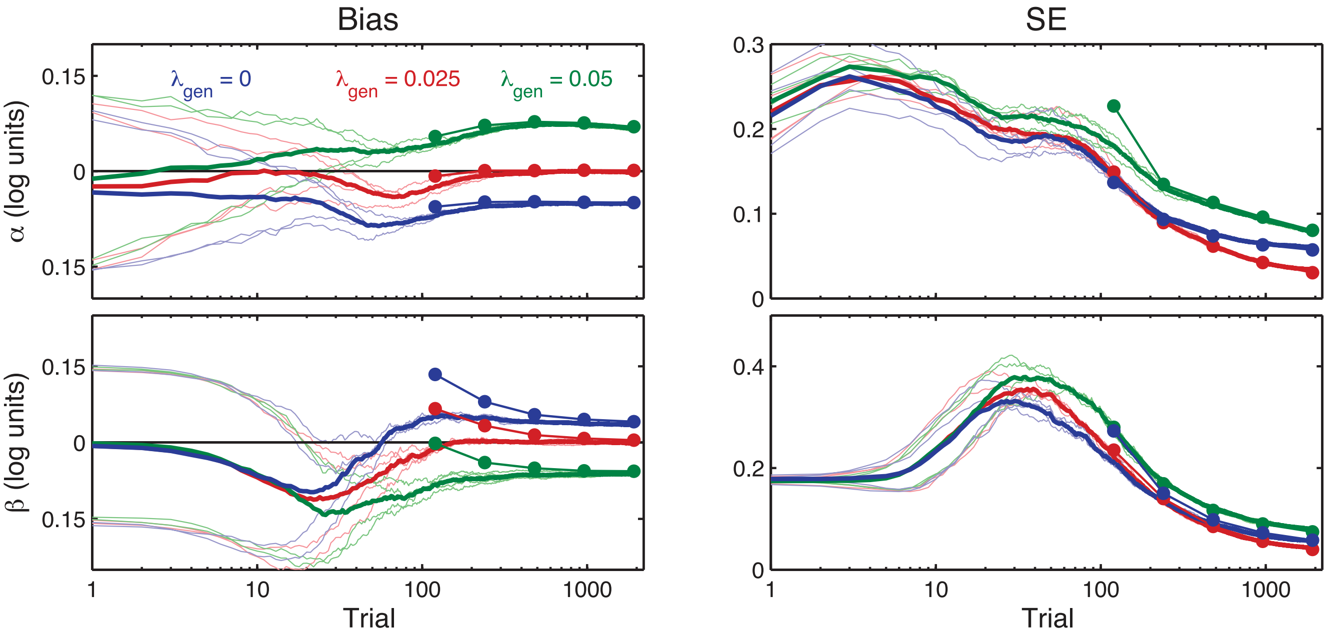

Figure 3. The effect of a mismatch between assumed and generating lapse rate on bias and standard error in Bayesian (lines) and ML (symbol) parameter estimates for data collected using the original psi-method. Thick lines show results for Bayesian estimates averaged across all 2,000 simulations in each condition. Thin lines show results separated out by the four sets of priors shown in Figure 2. Note that thin lines are color-coded by generating lapse rate (i.e., color codes do not correspond to colors used in Figure 2). Convergence of the results for the four different priors (which happens generally around Trial 100 or earlier) can be taken to indicate that from this point forward results are no longer affected by the exact choice of limits on the prior.

than the slope. Low N datasets are often consistent with extremely steep or shallow slopes and this will initially increase SE. As was found by Kontsevich and Tyler, only after about 30 trials or so does the psi-method begin to address the slope and only at this point does SE of slope begin to decrease.

Figures 4a and 5a display Bayesian and ML parameter estimates respectively for all simulations in which the generating lapse rate was 0.025 in the form of scatterplots of threshold and slope estimates. Also shown in these figures are the distributions of parameter estimates in the form of histograms. At all values of N, threshold and slope estimates are positively correlated (cf. Figure 1c).

Simulations refitted while lapse rate is allowed to vary

It is possible, of course, to free the lapse rate while fitting data even if the psi-method that was used to collect these data assumed a fixed value for the lapse rate. Thus, parameter estimates for the data obtained in the above simulations were also derived while allowing the lapse rate to vary. Bayesian estimates were derived by creating a posterior distribution across threshold, slope, and lapse rates. For each simulation, the same prior limits on threshold and slope values were used that were used during the psi-method itself. However, the prior now was 3D: It also contained different values for the lapse rate. Lapse rate values included in the posterior were limited to the interval [0, 0.1] in steps of 0.01. Bayesian estimates were derived again as the marginal means in the posterior distribution, now for all three of the parameters included in the posterior. ML estimates were also derived while allowing the lapse rate to vary. The value for the lapse rate was not constrained for the ML fits (other than to lie within the interval [0, 1]). Bias and standard errors of estimates are displayed in Figure 6, while scatterplots are shown in Figures 4b and 5b. Threshold and slope estimates appear to asymptote towards their true generating value (as they must eventually, e.g., Edwards, 1972). However, bias in parameter estimates is still substantial even at the highest number of trials used here. The mechanism behind the bias was discussed at some length in Prins (2012a). Briefly, bias arises because few observations are made at high values of F and, as a result, the data are fit about equally well by a family of functions which have different values of threshold, slope, and lapse rate (cf. Figure 1d). In other words, the posterior density is elongated with respect to the lapse rate dimension (see Figure 1a). Note from Figure 6 that the Bayesian lapse rate estimate starts to deviate from the mean of the prior distribution (i.e., 0.05) only after a relatively large number of trials have been collected and even then changes only at a very slow rate. Bayesian and ML estimates fail to converge even at $N =$

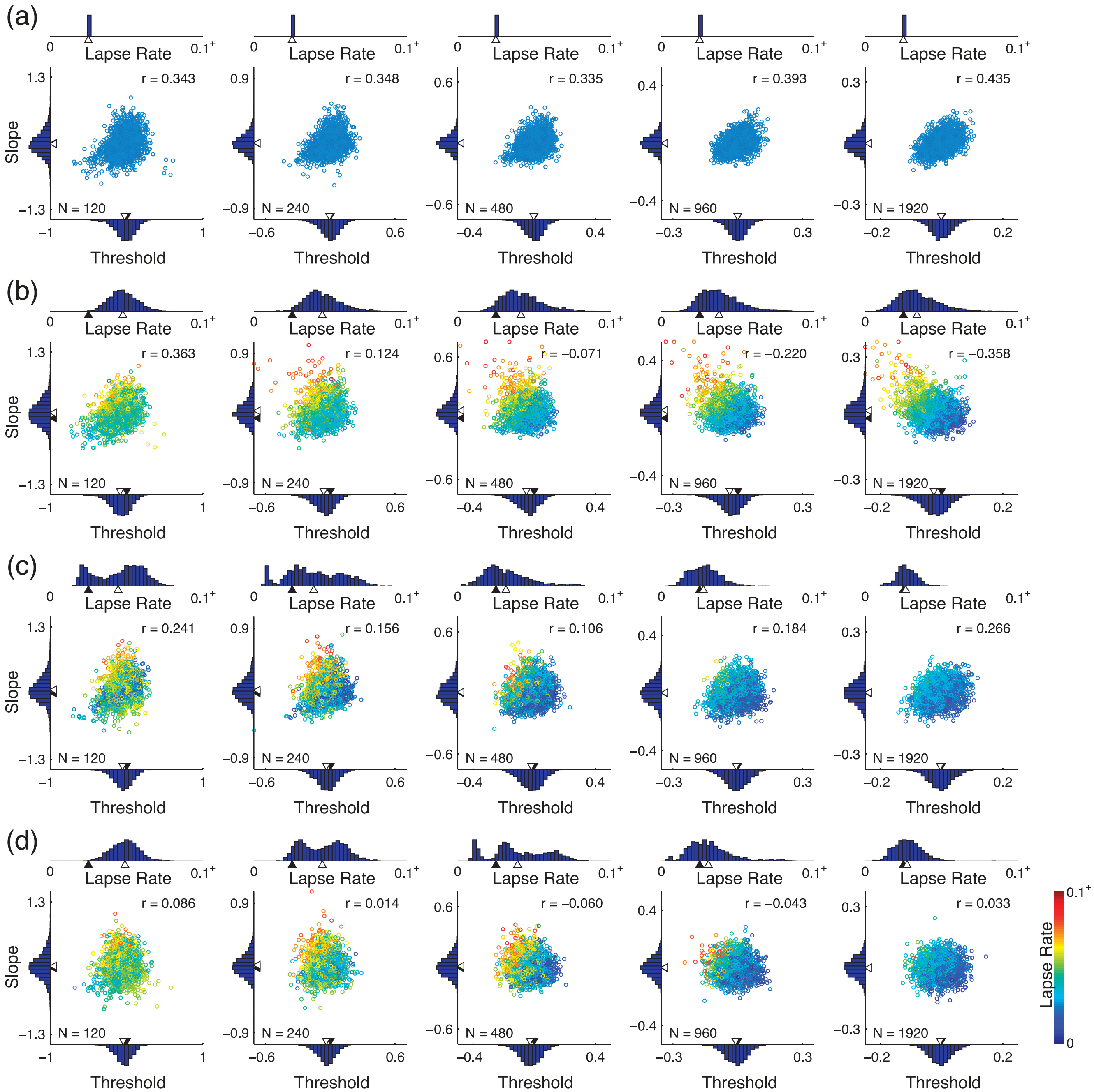

Figure 4. Bayesian parameter estimates for (a) original psi-method assuming a fixed lapse rate, (b) same data refitted with lapse rate free to vary, (c) psi$_{\alpha\beta\lambda}$ (psi^{+}) method, and (d) psi$_{\alpha\beta(\lambda)}$ method. Generating lapse rate was 0.025 for all. Filled triangular symbols indicate generating values of parameters; open triangular symbols indicate the mean of the estimates from all 2,000 simulations. Where only the open triangular symbol is visible, it obscures the filled symbol. Marginal threshold and slope distributions are shown as histograms on axes of the scatterplots. Lapse rate distributions are shown as histograms in separate plot. Note that data here are trimmed in that parameter estimates that exceed the limits of axes are assigned the value of this limit. Note that this was only done for graphical purposes here: Means of parameter estimates reported (open triangular symbols as well as the means and *SE*s presented in Figure 3 [and Figures 6 and 7]) are based on results that were not trimmed.

1,920 trials. This is primarily because very little information regarding the value of the lapse rate is available in the data. As a result, the Bayesian lapse rate estimates remain near the mean of the prior distribution on the lapse rate (i.e., 0.05; see the lapse rate histograms in Figure 4b). The ML lapse rate estimates tend to equal 0 (Figure 5b). Due to the dependency between lapse rate estimates and estimates

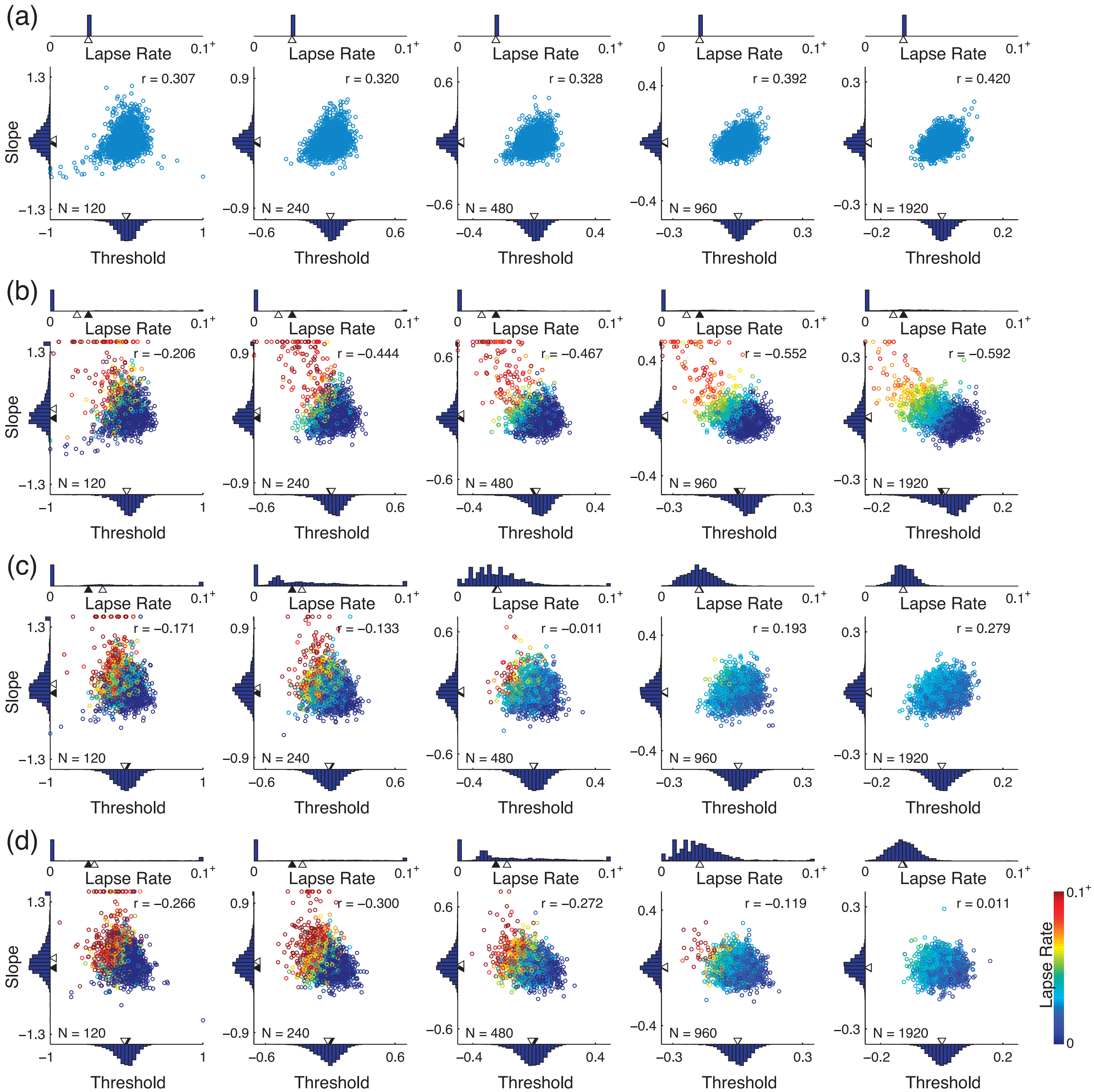

Figure 5. Same as Figure 4 but shown here are ML estimates.

for the threshold and slope parameters, systematic differences between ML and Bayesian lapse rate estimates are associated with systematic differences in threshold and slope estimates as well. Results for simulations using different values for the guess rate was equal to 0 and 0.25 are included in the Supplementary materials.

Experiment 2: Including the lapse rate in the posterior distribution: psi+ and psi-marginal methods

The observation was made above that the likelihood function and the Bayesian posterior are elongated in the lapse rate dimension when stimulus

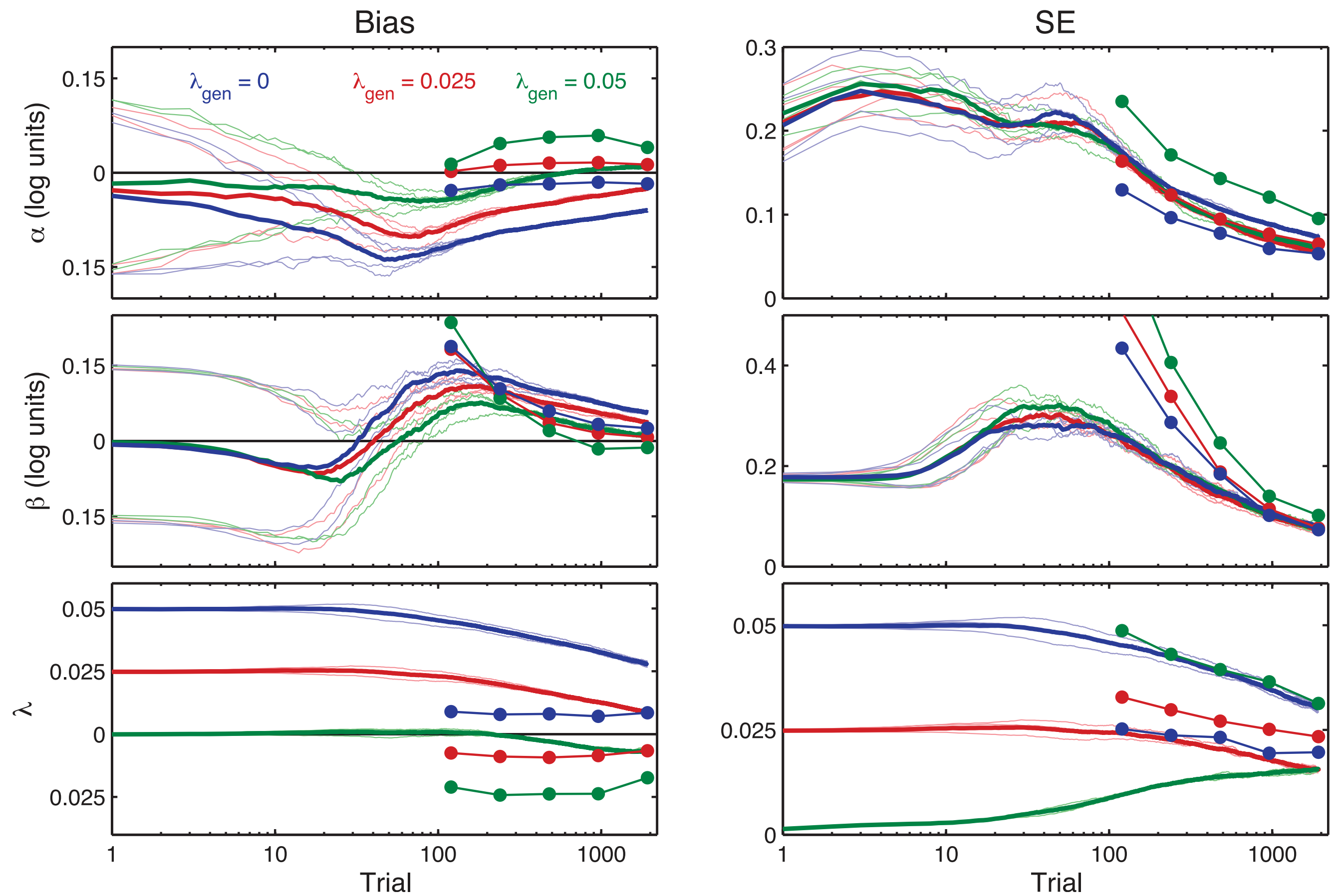

Figure 6. Bias and standard error of Bayesian (lines) and ML (symbols) parameter estimates when the data from Figure 3 are refitted while the lapse rate is allowed to vary.

collection is governed by the psi-method and that this elongation plays a large role in the bias observed in parameter estimates. An obvious solution to the observed bias then is to modify the psi-method to select stimulus intensities in order to minimize entropy in a posterior which includes the lapse rate. I have referred to such a strategy as the psi$^+$-method (Prins, 2012b). An obvious drawback of this strategy is that reducing uncertainty in the lapse rate parameter is made to be an explicit goal of the method. However, estimation of the lapse rate, being a nuisance parameter, should play a subservient role only. That is, our method should gather information regarding the lapse rate only insofar as this is the optimal strategy to reduce uncertainty in our parameters of primary interest. A second strategy then is to keep track again of a three-dimensional posterior distribution but to select stimulus intensities that will minimize expected entropy in the *marginal threshold x slope posterior distribution*. The marginal threshold × slope posterior distribution is derived simply by summation across the lapse rate dimension:

$$p_{\alpha,\beta}(\alpha = a,\ \beta = b) = \sum_{l} p(\alpha = a,\ \beta = b,\ \lambda = l), \qquad (4)$$

where $p(\alpha = a,\ \beta = b,\ \lambda = l)$ is the full posterior distribution defined across the threshold values a, slope values b and lapse rate values l that are contained in the parameter space. Consider again Figure 1. Figure 1(a) shows the full posterior distribution. The top panel under (b) is the marginal threshold × slope posterior derived by Equation 4. One may loosely conceive of this marginal distribution as the projection of the full posterior in (a) unto a threshold × slope plane (as if one were to look straight down on [a]). Since the high-density, cigar-shaped region is tilted, the marginal posterior displays more uncertainty regarding the threshold and slope values compared to a posterior that assumes a specific value for the lapse rate (i.e., one of the horizontal slices through the posterior shown in Figure 1c). In other words, the marginal threshold × slope distribution has incorporated the uncertainty regarding the value of the lapse rate. Since threshold, slope, and lapse rates show some degree of dependency, uncertainty with regard to the lapse rate contributes to the uncertainty

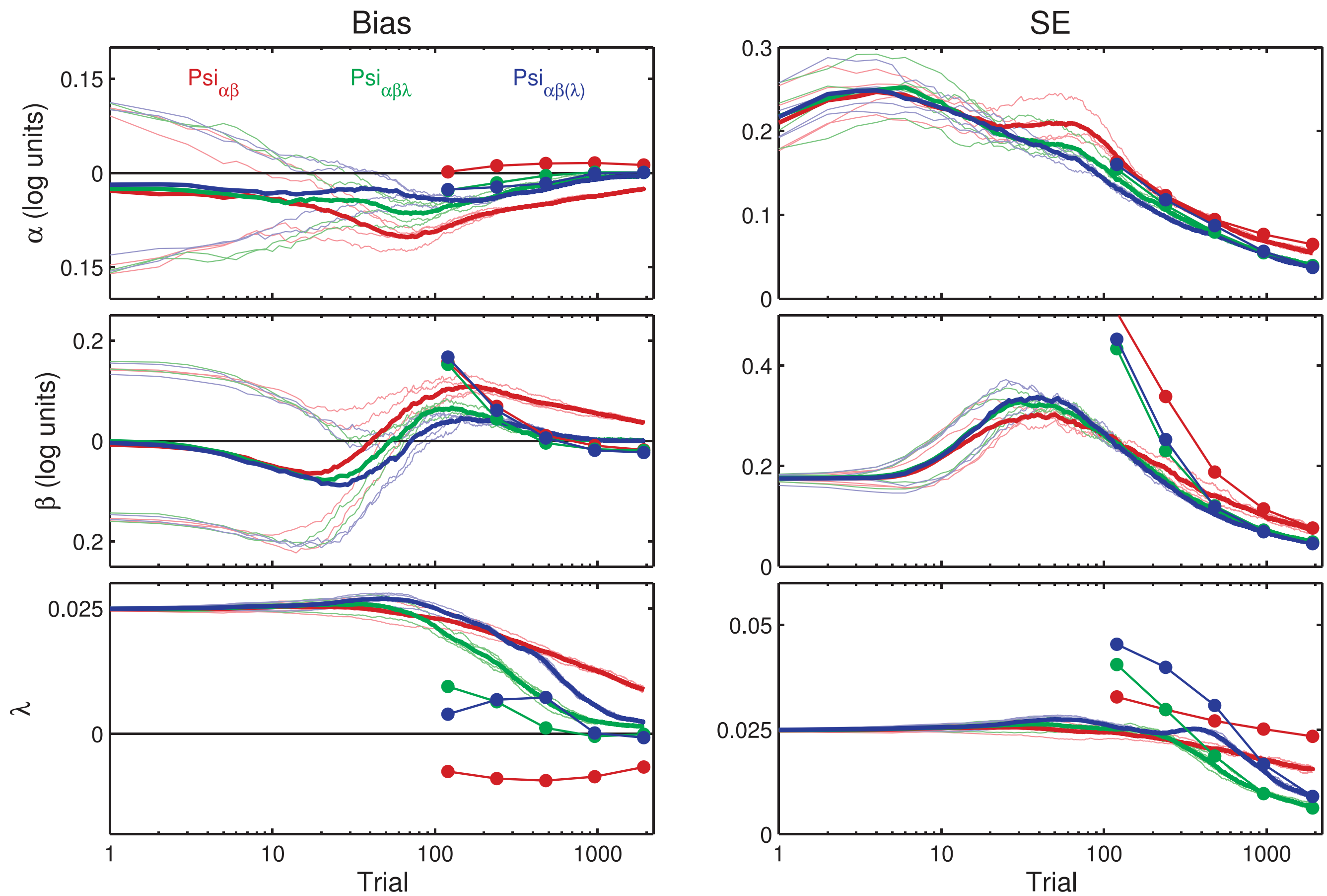

Figure 7. Bias and standard error of Bayesian (lines) and ML (symbols) parameter estimates for the original psi-method (i.e., $\mathrm{psi}_{\alpha\beta}$; replotted from Figure 6), $\mathrm{psi}_{\alpha\beta\lambda}$ method, and $\mathrm{psi}_{\alpha\beta(\lambda)}$ method.

in threshold and slope parameters. Thus, reduction of the entropy in the threshold × slope marginal posterior can be accomplished by gathering information regarding the value of the lapse rate. The appeal of this strategy is that all stimulus placements are motivated entirely by the criterion to maximize expected information gain regarding the values of the threshold and slope parameters only. In the process, the method may place stimuli at intensities at which information is gained primarily regarding the lapse rate, but only if that is the best placement to reduce uncertainty regarding the threshold and slope parameters. I refer to the general strategy of marginalizing nuisance parameters as the "psi-marginal method" and will use the shorthand notation $\mathrm{psi}_{\alpha\beta(\lambda)}$ to specify the particular implementation mentioned above. In this shorthand notation, the subscripted symbols indicate which parameters are included in the posterior distribution and the parentheses indicate the marginalization of a parameter. In this shorthand notation, "$\mathrm{psi}_{\alpha\beta}$" would correspond to the original psi-method and "$\mathrm{psi}_{\alpha\beta\lambda}$" would correspond to the psi^{+} method.

Methods

In order to investigate the behavior of the $\mathrm{psi}_{\alpha\beta\lambda}$ and the marginal $\mathrm{psi}_{\alpha\beta(\lambda)}$ methods, additional simulations were run. The procedures were identical to that described above for the psi-method, except for the following. For both the $\mathrm{psi}_{\alpha\beta\lambda}$ and the $\mathrm{psi}_{\alpha\beta(\lambda)}$ method a range of values for the lapse rate was included in the posterior distribution. Values for the lapse rate that were included ranged from 0 to 0.1, spaced in intervals of 0.01. In the $\mathrm{psi}_{\alpha\beta\lambda}$ method, after each trial the stimulus intensity associated with the lowest expected entropy in the full 3-D posterior was selected for the next trial. In the $\mathrm{psi}_{\alpha\beta(\lambda)}$ method, the stimulus intensity associated with the lowest expected entropy in the marginal posterior across threshold and slope values was selected for the next trial. In very loose terminology, whereas the goal of the $\mathrm{psi}_{\alpha\beta\lambda}$ method is to shrink

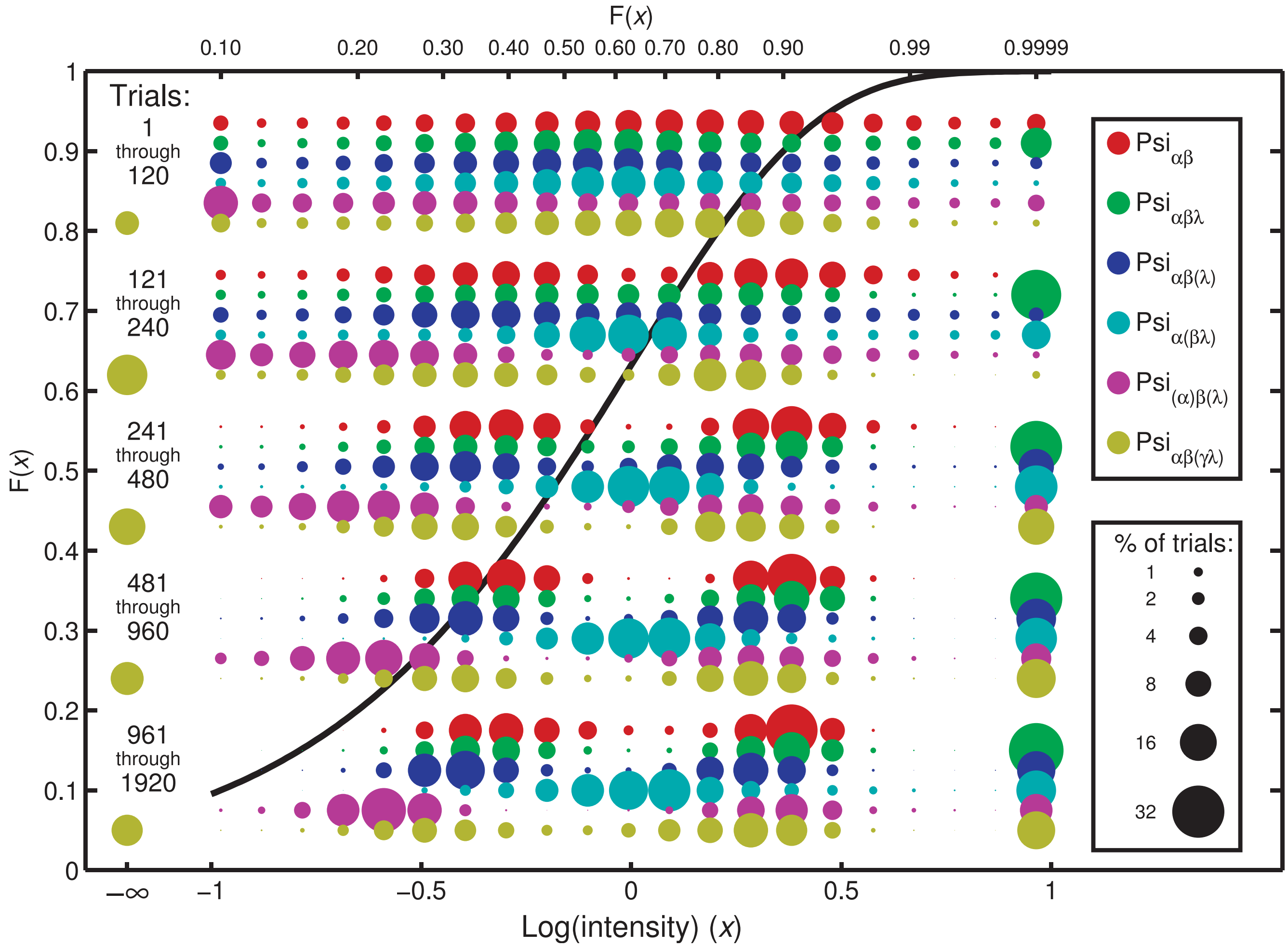

Figure 8. Average placement across stimulus intensities for the psi method (psi$_{\alpha\beta}$), psi^{+} method (psi$_{\alpha\beta\lambda}$), and various forms of the psi-marginal method. The area of the symbols is proportional to the number of trials presented at each intensity. Also shown is the generating function *F*.

the 3-D high-density "blob" in the posterior (see Figure 1a) in all three dimensions, the goal of the psi$_{\alpha\beta(\lambda)}$ method is to shrink the blob's projection unto the threshold $\times$ slope plane (see Figure 1b).

Results

Bias and *SE*s are shown in Figure 7, while relative placement of stimuli averaged across simulations is shown in Figure 8. Data shown for the original psi method (psi$_{\alpha\beta}$) in Figure 7 were also shown in Figure 3 and are included here again in order to allow easy comparison. Scatterplots of parameter estimates are shown in Figures 4c and 5c for the psi$_{\alpha\beta\lambda}$ method and in Figures 4d and 5d for the psi$_{\alpha\beta(\lambda)}$ method. Results indicate that Bayesian estimates of threshold and slope parameters asymptote towards their true values much faster when data are collected using the psi$_{\alpha\beta\lambda}$ and psi$_{\alpha\beta(\lambda)}$ methods compared to data collected using the traditional psi-method. The psi$_{\alpha\beta\lambda}$ and psi$_{\alpha\beta(\lambda)}$ methods perform about equally well, with the latter having a slightly lower SE in threshold estimate. Not surprisingly, both psi$_{\alpha\beta\lambda}$ and psi$_{\alpha\beta(\lambda)}$ lead to much more accurate and precise estimates of the lapse rate compared to psi$_{\alpha\beta}$. ML estimates and Bayesian estimates appear to converge at high N for data collected using psi$_{\alpha\beta\lambda}$ and psi$_{\alpha\beta(\lambda)}$ methods while they do not for data collected using the psi$_{\alpha\beta}$ method.

Replicating Kontsevich and Tyler (1999), the original psi-method starts off a run by targeting performance levels near the current threshold estimate. From the pattern of standard errors shown in Figure 3, it can be seen that the psi-method starts to target the slope estimate only after roughly 30 trials. From Figure 7, the psi$_{\alpha\beta\lambda}$ and psi$_{\alpha\beta(\lambda)}$ methods appear to do the same. The behavior of the psi$_{\alpha\beta\lambda}$ and psi$_{\alpha\beta(\lambda)}$ methods with regard to targeting of the lapse rate is most clearly seen in Figure 8. Not surprisingly, psi$_{\alpha\beta\lambda}$ places many trials at the highest stimulus intensity available to it even at $N =$

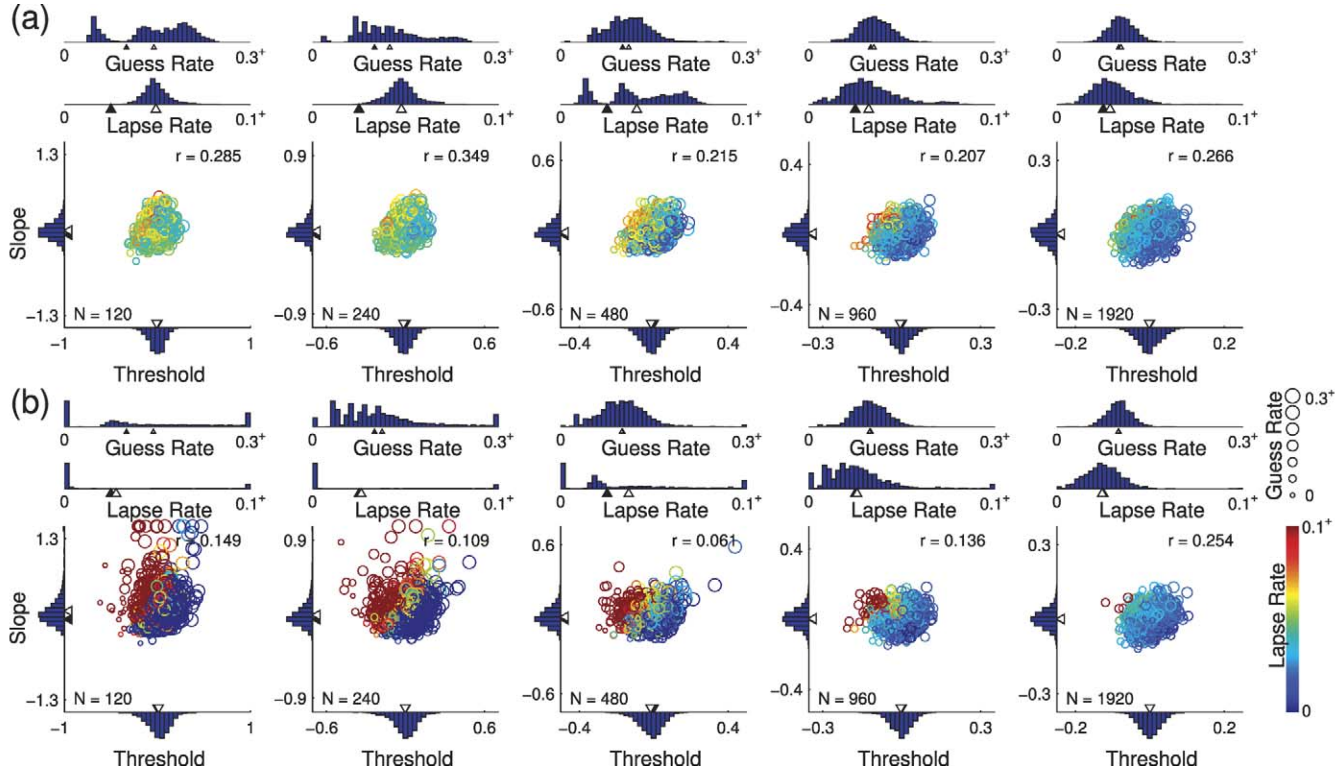

Figure 9. Bayesian (a) and ML (b) parameter estimates obtained using the psi$_{\alpha\beta(\gamma\lambda)}$ method in a yes/no task. Filled triangular symbols indicate generating values of parameters; open triangular symbols indicate the mean of the estimates from all 2,000 simulations. Where only the open triangular symbol is visible, it obscures the filled symbol. For graphical purposes, data are trimmed in the same manner as in Figure 4.

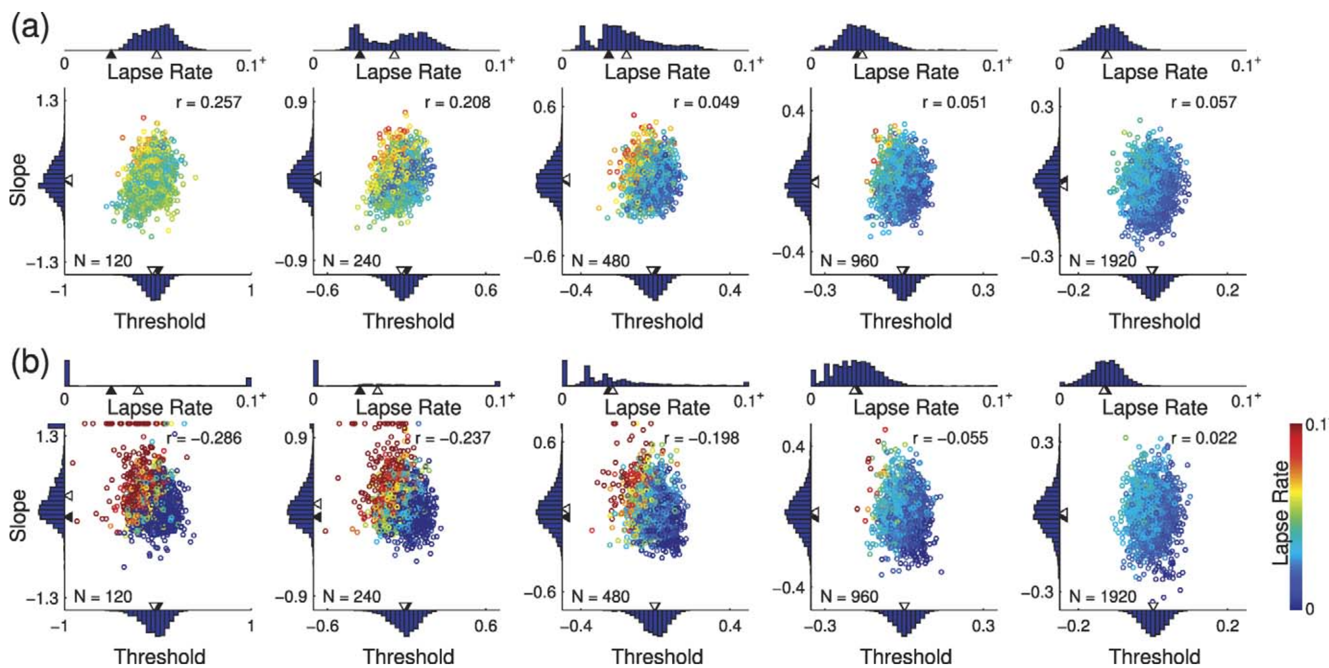

Figure 10. Bayesian (a) and ML (b) parameter estimates obtained using the psi$_{\alpha(\beta\lambda)}$ method in which threshold, slope and lapse rate are included in the prior but stimulus selection occurs such as to minimize entropy in the posterior after the slope and the lapse rate have been marginalized. Filled triangular symbols indicate generating values of parameters; open triangular symbols indicate the mean of the estimates from all 2,000 simulations. Where only the open triangular symbol is visible, it obscures the filled symbol. For graphical purposes, data are trimmed in the same manner as in Figure 4.

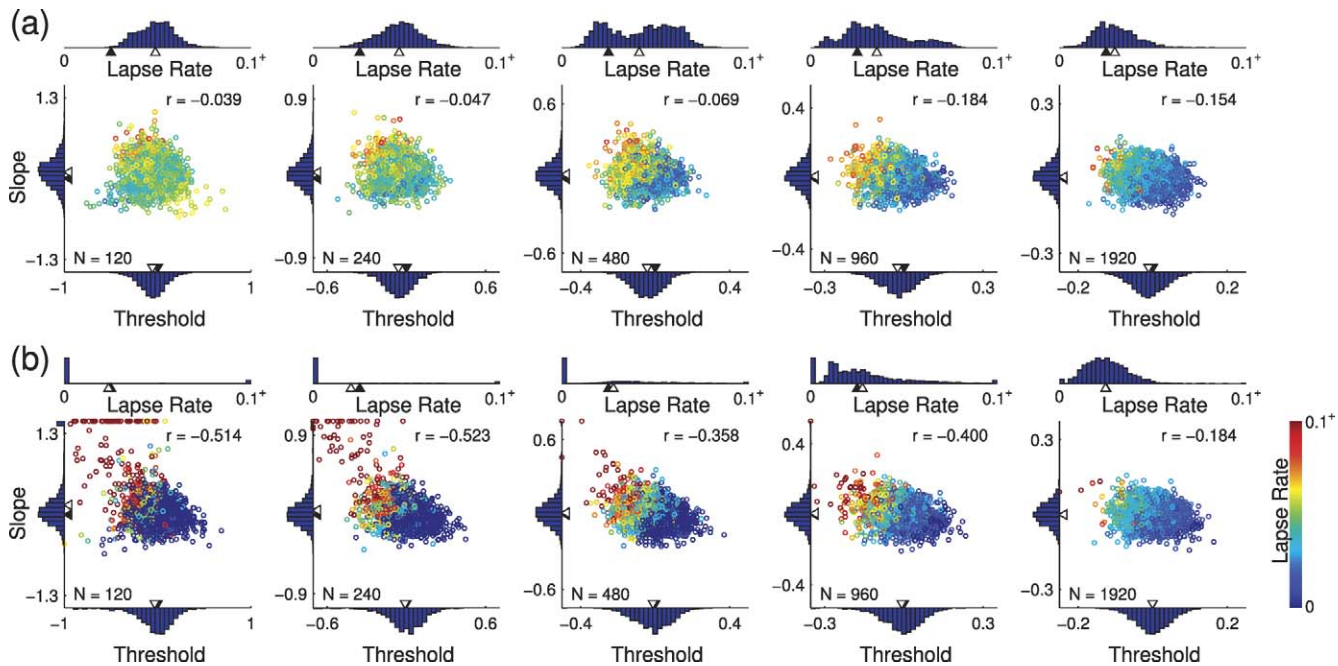

Figure 11. Bayesian (a) and ML (b) parameter estimates obtained using the psi$_{(\alpha)\beta(\lambda)}$ method in which threshold, slope, and lapse rate are included in the prior but stimulus selection occurs such as to minimize entropy in the posterior after the threshold and the lapse rate have been marginalized. The first 30 trials were conducted as psi$_{\alpha\beta(\lambda)}$ (see text for details). Filled triangular symbols indicate generating values of parameters; open triangular symbols indicate the mean of the estimates from all 2,000 simulations. Where only the open triangular symbol is visible, it obscures the filled symbol. Data are trimmed in the same manner as in Figure 5.

120: Reducing uncertainty in the lapse rate is an explicit goal of psi$_{\alpha\beta\lambda}$. While reducing uncertainty in the lapse rate is not an explicit goal of psi$_{\alpha\beta(\lambda)}$, it places a relatively large number of stimuli at the highest possible intensity also (though far fewer than psi$_{\alpha\beta\lambda}$ does). Note, however, that this is not true at low N. Apparently, at low N, including stimulus placements at very high intensities is not the most efficient manner in which to reduce uncertainty in threshold and slope parameter values. Note also that even though psi$_{\alpha\beta(\lambda)}$ does not place more stimuli at high intensity than the original psi-method does (fewer even) at $N = 120$ and $N = 240$, it far outperforms the psi-method in terms of bias and *SE* of the threshold and slope parameters.

Some aspects of the placement patterns in Figure 8 may seem erratic. For example, from Figure 8 and as noted above, psi$_{\alpha\beta}$ places more stimuli at the highest intensity during the first 120 trials than psi$_{\alpha\beta(\lambda)}$ does. At first glance, this seems counterintuitive. After all, psi$_{\alpha\beta}$ only has uncertainty regarding the threshold and the slope and very little information regarding either of these parameters is to be gained at this highest intensity. However, we need to realize that such placement is only unexpected to one who knows that the generating PF is near its asymptotic value at this intensity. The adaptive method does not know this. In fact, both psi$_{\alpha\beta}$ and psi$_{\alpha\beta(\lambda)}$ only place stimuli at the highest intensity this early in a run when responses thus far collected are consistent with a PF that has a very shallow slope. In other words, either method is under the impression (if you will) that it is placing a stimulus at a much lower intensity than it in fact is. Similarly, the apparent greater precision of placements of psi$_{\alpha\beta}$ compared to the other methods is not a result of strategy, rather it results from psi$_{\alpha\beta}$ having more information regarding the generating PF compared to the other methods (i.e., it knows the value of the lapse rate).

Other possibilities

Any combination of the four parameters of the PF may be included in the posterior and any of them may be marginalized (as long as at least one of the parameters is not marginalized). I will briefly discuss a few other possibilities. In a "yes/no" task the lower asymptote of the PF corresponds to the false alarm rate, which will generally be considered a nuisance parameter of unknown value. Thus, extending our logic above, we can include the guess rate in the posterior distribution but select placement such as to maximize the expected information gain in a posterior in which it has been marginalized. If we again consider the lapse rate to be a nuisance parameter also, the appropriate

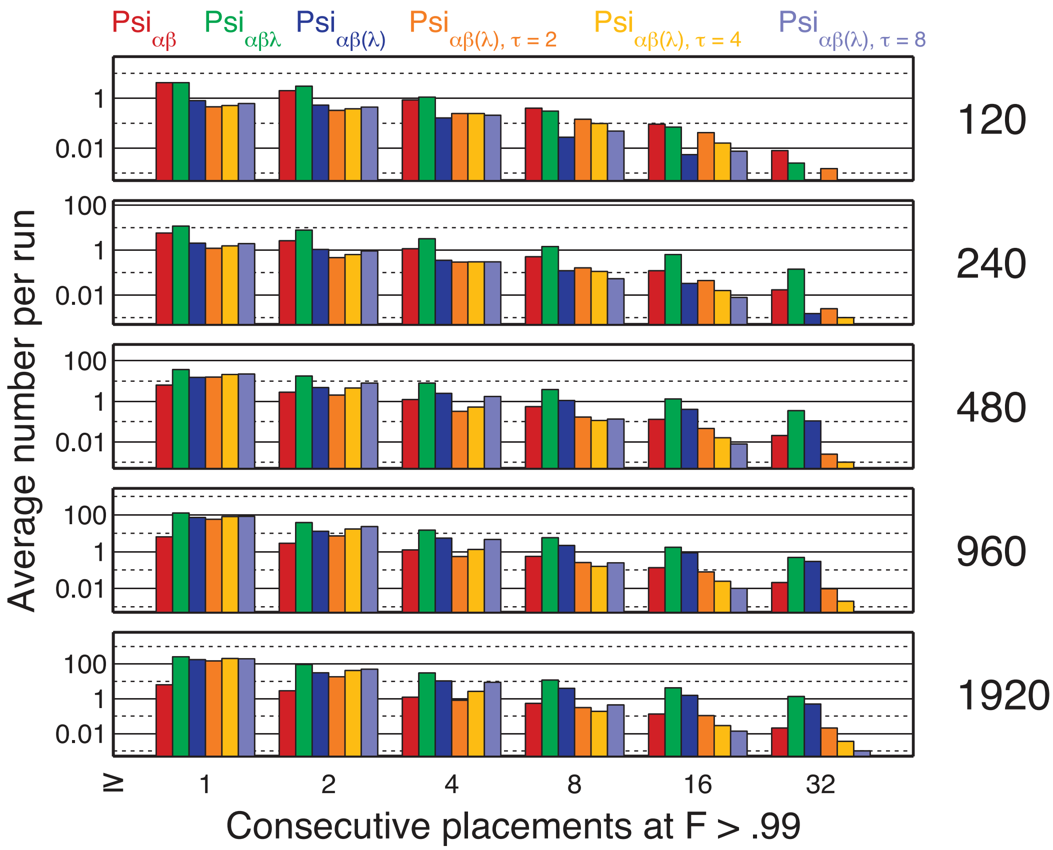

Figure 12. Average frequency of occurrence of series of consecutive stimulus placements at an intensity x at which the (generating) F (Equation 1) evaluates to $>$ 0.99. Note that results show frequency of series equal to or longer than the number listed on the abscissa. Results are shown for the psi$_{\alpha\beta}$, psi$_{\alpha\beta\lambda}$, and psi$_{\alpha\beta(\lambda)}$ methods. For the latter, results are also shown when the psi$_{\alpha\beta(\lambda)}$ selection criterion was temporarily suspended in favor of psi$_{\alpha\beta}$ (i.e., assume fixed lapse rate) for three values of Wait Time τ (see text for detail).

method would be psi$_{\alpha\beta(\gamma\lambda)}$. By way of demonstration, 2,000 simulations were run which were identical to the psi$_{\alpha\beta(\lambda)}$ runs above except for the following. The generating guess rate was arbitrarily set to 0.1. Guess rates were included in the posterior, which was now 4-D (threshold $\times$ slope $\times$ guess rate $\times$ lapse rate). The values for the guess rates included were 0 through 0.3 in steps of 0.03. The values included for the other parameters were as above. A stimulus of zero intensity (i.e., $\log(x) = -\infty$) was added to the 21 stimulus intensities described above. After each trial, the stimulus intensity to be used on the next trial was selected based on the criterion that it minimizes the expected entropy in the marginal threshold $\times$ slope posterior. Results are shown in Figure 9.

Another possibility I highlight here is psi$_{\alpha(\beta\lambda)}$ which maintains a posterior distribution across threshold, slope and lapse rates (as above) but selects stimulus intensities such as to optimize estimation of the threshold only. By way of demonstration, 2,000 simulations were performed exactly as those performed in Experiment 2 except that now only the threshold was regarded as a parameter of primary interest while the slope and lapse rate were regarded as nuisance parameters by marginalizing them. Resulting Bayesian and ML estimates are shown in Figure 10. From Figure 10, it is clear that the method has adjusted its placement and that this has resulted in a higher precision of the threshold estimate while precision in the slope has decreased.

Finally, I consider here psi$_{(\alpha)\beta(\lambda)}$. In words, only the slope parameter is considered to be of interest, the threshold and slope are considered to be nuisance parameters. Note that while it is common practice to fix the slope when one is interested in the threshold (e.g., Pentland's [1980] Best PEST and Watson & Pelli's [1983] Quest do this), fixing the value of the threshold when one is interested in the slope would be a grave error indeed. Again, 2,000 simulations were performed exactly as those performed in Experiment 2 except that the threshold and the lapse rate were regarded as nuisance parameters. As it turns out, this condition benefits greatly from having some information regarding the location of the PF at the beginning of the run. For that reason, the threshold parameter was marginalized only after 30 trials. That is, for Trials 1 through

30 the method was psi$_{\alpha\beta(\lambda)}$, for the remaining trials it was psi$_{(\alpha)\beta(\lambda)}$. Note that any of the parameters in the posterior may be marginalized or demarginalized at any point during a run. For example, another possibility would have been to start the run off as psi$_{\alpha(\beta\lambda)}$, then switch to psi$_{(\alpha)\beta(\lambda)}$. Resulting Bayesian and ML estimates are shown in Figure 11, from which it is clear that the slope estimate has benefitted from being the only parameter of primary interest while the estimate of the threshold, being defined as a nuisance parameter, has become less precise.

As in the original psi-method, the prior distribution on any of the parameters of the PF need not be (constrained) uniform as they were in all of the simulations above. Researchers may specify any form of prior probability distribution on any of the parameters included in the prior. This includes those parameters designated nuisance parameters by marginalizing them. Broadly speaking, the more specific the prior on a parameter is, the less information the psi-method will attempt to gain about the parameter. Below it is shown how this property may be used to avoid lengthy consecutive placements at asymptotic performance levels.

Limitations

By their nature, the psi-method and its variations proposed here are so-called "greedy" algorithms in that they consider only one trial ahead. King-Smith, Grigsby, Vingrys, Benes, and Supowit (1994) investigated the advantage of looking two trials ahead compared to looking one trial ahead in a procedure similar to the psi-method and found that there was little advantage to looking two trials ahead. There is no reason to suspect that this is different for the proposed method. The method has looked only one trial ahead in all of the simulations above and it compared favorably to the psi-method.

When the lapse rate is included in the posterior distribution occasionally relatively long consecutive series of high-intensity stimulus placements may occur (Figure 12). For the simulations performed in Experiment 2 and that were controlled by the psi$_{\alpha\beta(\lambda)}$ method this is not an issue at values of N up to 240. As a matter of fact, at $N = 120$, psi$_{\alpha\beta(\lambda)}$ places *fewer* trials near upper asymptotic levels of performance compared to the original psi-method (Figure 8) and also has fewer lengthy consecutive series of near asymptotic performance placements. At larger N, however, lengthy series of consecutive placements at near-asymptotic performance tend to occur. For example, in the psi$_{\alpha\beta(\lambda)}$ simulations, a run of 480 trials contained on average about one series of eight or more consecutive stimulus placements at the highest intensity available to the method (Figure 12). Such lengthy series of "free trials" will almost certainly affect an observer's level of vigilance. A few strategies may be used to avoid such consecutive series of high-intensity placements. For example, the problem will be lessened when multiple conditions, each controlled by its own psi-marginal run, are randomly interleaved in a testing session. Any series of consecutive high intensity placements that may occur in any given run will then be interrupted by trials from the other runs.

Another manner in which to avoid consecutive placements at high intensities is to change the optimization criterion temporarily. Consider again the simulations in Experiment 2 in which the posterior was defined across threshold, slope, and lapse rate values and the selection criterion was based on minimizing entropy in the marginal threshold × slope posterior. One may at any time temporarily change the selection criterion such that the method minimizes expected entropy within a constant-lambda slice through the posterior (i.e., as the original psi-method does). This selection criterion will generally not result in the selection of a stimulus intensity that corresponds to near-asymptotic performance (though it may, especially early on in a run). This strategy may be used to avoid consecutive placements near asymptotic performance. As a demonstration, simulations were run that were identical to those controlled by the psi$_{\alpha\beta(\lambda)}$ method in Experiment 2, except that after any trial in which placement was at the highest stimulus intensity available to the method, the selection criterion was temporarily changed to that of the original psi-method. The value at which the lapse rate was fixed on such trials corresponded to the current ML estimate of the lapse rate. After a "wait time" that was randomly drawn from an exponential mass function (so as to maintain "constant hazard") the method resorted back to selecting the stimulus intensity that maximized expected information gain in the marginal threshold × slope posterior. Three different mean wait times (τ) were used: two, four, or eight trials. The frequency of occurrence of consecutive placements at near-asymptotic levels for these simulations is included in Figure 12. Note that the strategy to suspend the psi-marginal selection criterion temporarily to favor that of the original psi-method did not affect bias or precision of threshold and slope estimates significantly (bias and standard errors shown in Supplementary materials).

Conclusions

A new, flexible adaptive method is proposed that is a variation of Kontsevich and Tyler's (1999) psi-method.

This psi-marginal method is the first stimulus selection method to deal with nuisance parameters in an adaptive manner. It does not require one to assume a fixed lapse rate, which was shown to result in bias in parameter estimates when the assumed and actual lapse rate do not match. In the context for which the original psi-method was developed (i.e., an mAFC task) it was shown to outperform the psi-method in bias and precision when even a moderate degree of mismatch between assumed and actual lapse rate exists.

The psi-marginal method is highly flexible. The method allows users to designate each of the four parameters of a PF individually to be either a parameter of primary interest whose estimation should be optimized, a nuisance parameter to be estimated only insofar it will benefit the efficiency with which the parameters of primary interest are estimated, or as a fixed parameter with known or assumed value. This paper focused on a common experimental procedure, the mAFC task. A few other examples were given here also. It was demonstrated here in the context of a "yes/no" task. Results indicated that estimates of both threshold and slope were virtually free of bias in this task also. It was also shown to give higher precision in the threshold estimate when the threshold is designated to be the only parameter of primary interest and the slope and the lapse rate are treated as nuisance parameters. Finally, precision in the slope estimate was increased when it was designated as the only parameter of primary interest and the threshold as well as lapse rate were designated to be nuisance parameters.

The psi-marginal method is implemented in the Palamedes toolbox (Prins & Kingdom, 2009). Self-contained Matlab demonstration code is included in the Supplemental materials.

Keywords: psychometric function, Bayesian adaptive method, alternative forced choice, yes/no task, maximum likelihood

Supplementary materials

Supplementary materials include results for conditions not included in this manuscript and Matlab code that demonstrates various methods proposed here.

Acknowledgments

This work was supported by the Department of Psychology, University of Mississippi.

Commercial relationships: none.
Corresponding author: Nicolaas Prins.
Email: nprins@olemiss.edu.
Address: Department of Psychology, University of Mississippi, Oxford, MS, USA.

References

Dixon, W. J., & Mood, A. M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association, 43,* 109–126.

Edwards, A. W. F. (1972). *Likelihood*. Cambridge, UK: Cambridge University Press.

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. *Vision Research, 38,* 1861–1881.

Hall, J. L. (1981). Hybrid adaptive procedure for estimation of psychometric functions. *The Journal of the Acoustical Society of America, 69,* 1763–1769.

Harvey, L. O. (1986). Efficient estimation of sensory thresholds. *Behavior Research Methods, Instruments, & Computers, 18,* 623–632.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics, 49,* 227–229.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation, and practical implementation. *Vision Research, 34,* 885–912.

King-Smith, P. E., & Rose, D. (1997). Principles of an adaptive method for measuring the slope of the psychometric function. *Vision Research, 37,* 1595–1604.

Kingdom, F. A. A., & Prins, N. (2010). Psychophysics: A practical introduction. London, UK: Academic Press.

Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics, 63,* 1421–1455.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research, 39,* 2729–2737.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49,* 467–477.

Manny, R. E., & Klein, S. A. (1985). A three alternative tracking paradigm to measure Vernier acuity of older infants. *Vision Research, 25*(9), 1245–1252.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7,* 308–313.

Pentland, A. (1980). Maximum likelihood estimation: The best PEST. *Perception & Psychophysics, 28,* 377–379.

Prins, N. (2012a). The psychometric function: The lapse rate revisited. *Journal of Vision, 12*(6):25, 1–16, http://www.journalofvision.org/content/12/6/25, doi:10.1167/12.6.25. [PubMed] [Article].

Prins, N. (2012b). The adaptive psi method and the lapse rate. *Journal of Vision, 12*(9): 322, http://www.journalofvision.org/content/12/9/322, doi:10.1167/12.9.322. [Abstract].

Prins, N., & Kingdom, F. A. A. (2009). *Palamedes: Matlab® routines for analyzing psychophysical data*. Internet site: www.palamedestoolbox.org.

Swanson, W. H., & Birch, E. E. (1992). Extracting thresholds from noisy psychophysical data. *Perception & Psychophysics, 51*(5), 409–422.

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *The Journal of the Acoustical Society of America, 41,* 782–787.

Taylor, M. M., Forbes, S. M., & Creelman, C. D. (1983). PEST reduces bias in forced choice psychophysics. *The Journal of the Acoustical Society of America, 74*(5), 1367–1374.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research, 35*(17), 2503–2522.

Treutwein, B., & Strasburger, H. (1999). Fitting the psychometric function. *Perception & Psychophysics, 61*(1), 87–106.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33,* 113–120.

Watt, R. J., & Andrews, D. P. (1981). APE: Adaptive probit estimation of psychometric Functions. *Current Psychological Reviews, 1,* 205–214.

Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *The British Journal of Mathematical and Statistical Psychology, 18,* 1–10.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*(8), 1293–1313.