The psychometric function: The lapse rate revisited

Nicolaas Prins

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

In their influential paper, Wichmann and Hill (2001) have shown that the threshold and slope estimates of a psychometric function may be severely biased when it is assumed that the lapse rate equals zero but lapses do, in fact, occur. Based on a large number of simulated experiments, Wichmann and Hill claim that threshold and slope estimates are essentially unbiased when one allows the lapse rate to vary within a rectangular prior during the fitting procedure. Here, I replicate Wichmann and Hill's finding that significant bias in parameter estimates results when one assumes that the lapse rate equals zero but lapses do occur, but fail to replicate their finding that freeing the lapse rate eliminates this bias. Instead, I show that significant and systematic bias remains in both threshold and slope estimates even when one frees the lapse rate according to Wichmann and Hill's suggestion. I explain the mechanisms behind the bias and propose an alternative strategy to incorporate the lapse rate into psychometric function models, which does result in essentially unbiased parameter estimates.

Keywords: psychophysical methods, psychometric function, lapse rate, maximum-likelihood

Citation: Prins, N. (2012). 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.

Introduction

The psychometric function (PF) relates some behavioral measure (e.g., proportion correct on a detection task) to some quantitative characteristic of a sensory stimulus (e.g., luminance contrast). In the following I will refer to the latter simply as stimulus *intensity*, though this may not be an appropriate term in many circumstances (e.g., the variable may be spatial or temporal frequency, orientation offset, etc.). A generic formulation of the psychometric function is given by:

$$\psi(x;\alpha,\beta,\gamma,\lambda) = \gamma + (1-\gamma-\lambda)\mathrm{F}(x;\alpha,\beta) \qquad (1a)$$

(e.g., Wichmann & Hill, 2001, Kingdom & Prins, 2010). Though discredited, the classic high-threshold detection model (e.g., Swets, 1961) provides for an intuitively appealing interpretation of the parameters of Equation 1a. Under the high-threshold model, $\mathrm{F}(x;\ \alpha,\ \beta)$ describes the probability of detection by an underlying sensory mechanism as a function of stimulus intensity x, γ corresponds to the guess rate (the probability of a correct response when the stimulus is not detected by the underlying sensory mechanism), and λ corresponds to the lapse rate (the probability of an incorrect response, which is independent of stimulus intensity). Several forms of $\mathrm{F}(x;\ \alpha,\ \beta)$ are in common use such as the Logistic function, the Weibull function, and the cumulative normal distribution. In this paper, the Weibull function is used exclusively and is given by:

$$F_W(x;\alpha,\beta) = 1 - \exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right) \qquad (1b)$$

The parameter α of $\mathrm{F_W}(\mathrm{x};\ \alpha,\ \beta)$ determines the function's location and is commonly referred to as the function's 'threshold.' The parameter β determines the rate of change of performance as a function of stimulus intensity x and is commonly referred to as the 'slope.'

Even though the high-threshold model has been discredited (e.g., Swets, 1961), Equation 1 is consistent also with assumptions of signal-detection theory as proved formally by García-Pérez and Alcalá-Quintana (2007). Either way, few theorists would argue with the notion that whereas the threshold and slope parameters characterize the sensory mechanism that underlies performance, the remaining two parameters do not. Rather, the guess rate characterizes the decision process and the lapse rate characterizes such things as observer vigilance and response error. While the guess rate can generally be assumed to have a value determined by the experimental procedure (e.g., in an m-AFC task the guess rate can be assumed to equal 1/m), the same cannot be said of the lapse rate. As it does not describe the sensory mechanism, researchers generally are not interested in the value of the lapse rate per se.

While the effect that lapses have on parameter estimates has been noted for some time (for example, Manny & Klein addressed the issue as early as 1985), systematic investigations of this effect (e.g., Treutwein & Strasburger, 1999; Wichmann & Hill, 2001) are relatively recent. Wichmann and Hill (2001) advocate allowing the value of the lapse rate to vary alongside the values of the threshold and slope of the PF. They do so based on the results of a large number of simulated experiments. Briefly, Wichmann and Hill

doi: 10.1167/12.6.25 Received December 6, 2011; published June 19, 2012 ISSN 1534-7362 © 2012 ARVO

produced simulated datasets that were generated by a Weibull function with known parameter values. The generating values for α and β were equal to 10 and 3, respectively. The guess rate γ was 0.5. The generating lapse rate λ was systematically varied from 0 to 0.05 in steps of 0.01. The method of constant stimuli (MOCS) utilizing seven different stimulus placement regimens was used. The seven stimulus placement regimens are shown in Figure 1 (s1 through s7) relative to the generating form of F. The total number of simulated trials (N) in each simulated experiment was evenly distributed among the six stimulus intensities in each of the placement regimens. Each simulated dataset was then fitted with the psychometric function in Equation 1 using a maximum-likelihood criterion. The threshold and slope parameters were free to vary during the fitting process. The lapse rate parameter was either held constant at a fixed value or was allowed to vary within the interval [0 0.06]. This prior[1] was placed on the lapse rate parameter to reflect beliefs regarding likely values of the lapse rate parameter. Unless the prior is applied, nonsensical negative estimates of the lapse rate might result, as well as unrealistically high estimates of the lapse rate.

Wichmann and Hill (2001) report the threshold and slope not in terms of α and β, but rather in terms of $F^{-1}_{0.5}$ [that is, the stimulus intensity at which function F_W (Equation 1b) evaluates to 0.5] and $F'_{0.5}$ (that is, the gradient or first derivative of F evaluated at $F^{-1}_{0.5}$). The true, generating, values of these quantities are 8.85 and 0.118 respectively. Figure 2 was taken from Wichmann and Hill (2001). It shows the median threshold and median slope estimates, each derived based on 2,000 simulated experiments. The light symbols show the estimates when the lapse rate was fixed at zero, the darker symbols show the estimates when the lapse rate was allowed to vary. The different shapes of the symbols in the figure indicate stimulus placement regimen and correspond to the symbols used in Figure 1. The true (generating) values of the threshold (in terms of $F^{-1}_{0.5}$) and slope (in terms of $F'_{0.5}$) are indicated by the horizontal lines.

As is clear from Figure 2, both the threshold and slope estimates are significantly biased when the lapse rate is assumed to equal 0 but the generating lapse rate in fact differs from 0. The severity of the bias is (predictably) mainly a function of whether the stimulus placement regimen included stimulus placements at high intensities (see Wichmann & Hill, 2001 for a detailed argument on why this is so). However, when the lapse rate was allowed to vary, threshold and slope estimates were, in their terminology, “essentially unbiased” (Wichmann & Hill, 2001, p. 1298).

Attempted replication of Wichmann and Hill

García-Pérez and Alcalá-Quintana (2005) have noted that when during the fitting procedure the lapse rate estimate is allowed to vary but constrained to have

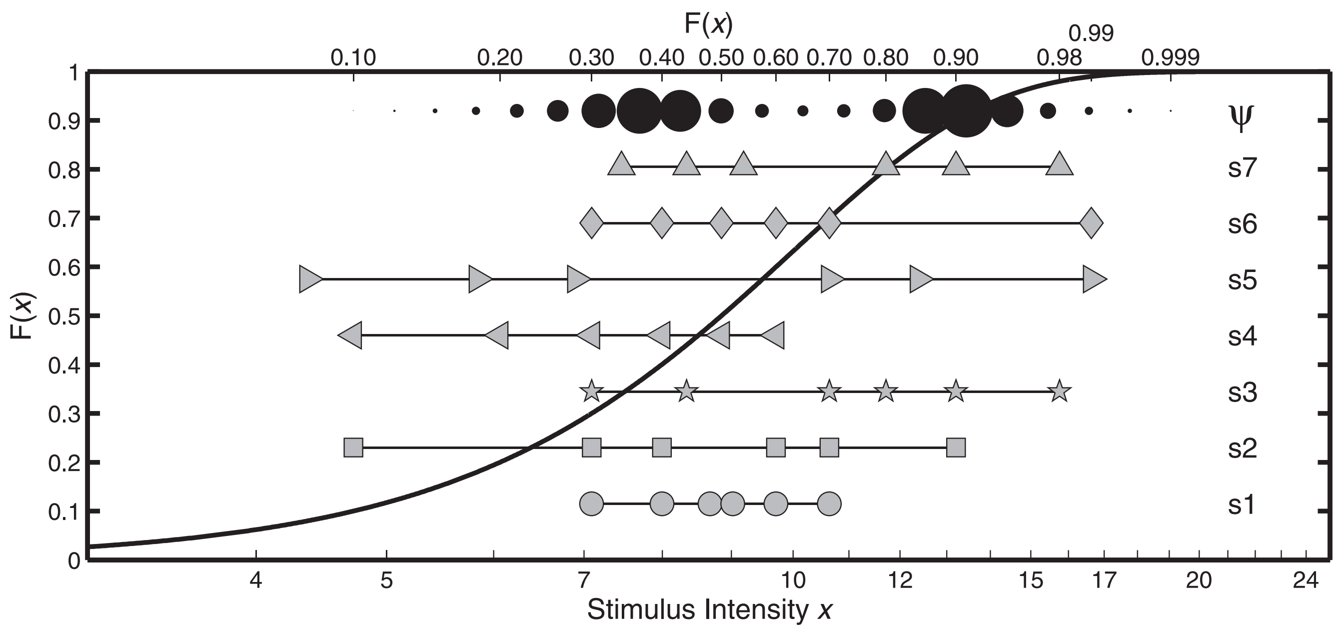

Figure 1. The seven different stimulus placement regimens (s1 through s7) used by Wichmann and Hill (2001). Also shown is the stimulus placement resulting from using the adaptive psi-method (Ψ) averaged across 10,000 simulated runs of 960 trials each (see Method section for more details). In the latter, the area of the symbols in the Figure is proportional to the number of trials presented at the corresponding stimulus intensity. The curve shown is the Weibull with $\alpha = 10$, $\beta = 3$ (i.e., the generating form of F).

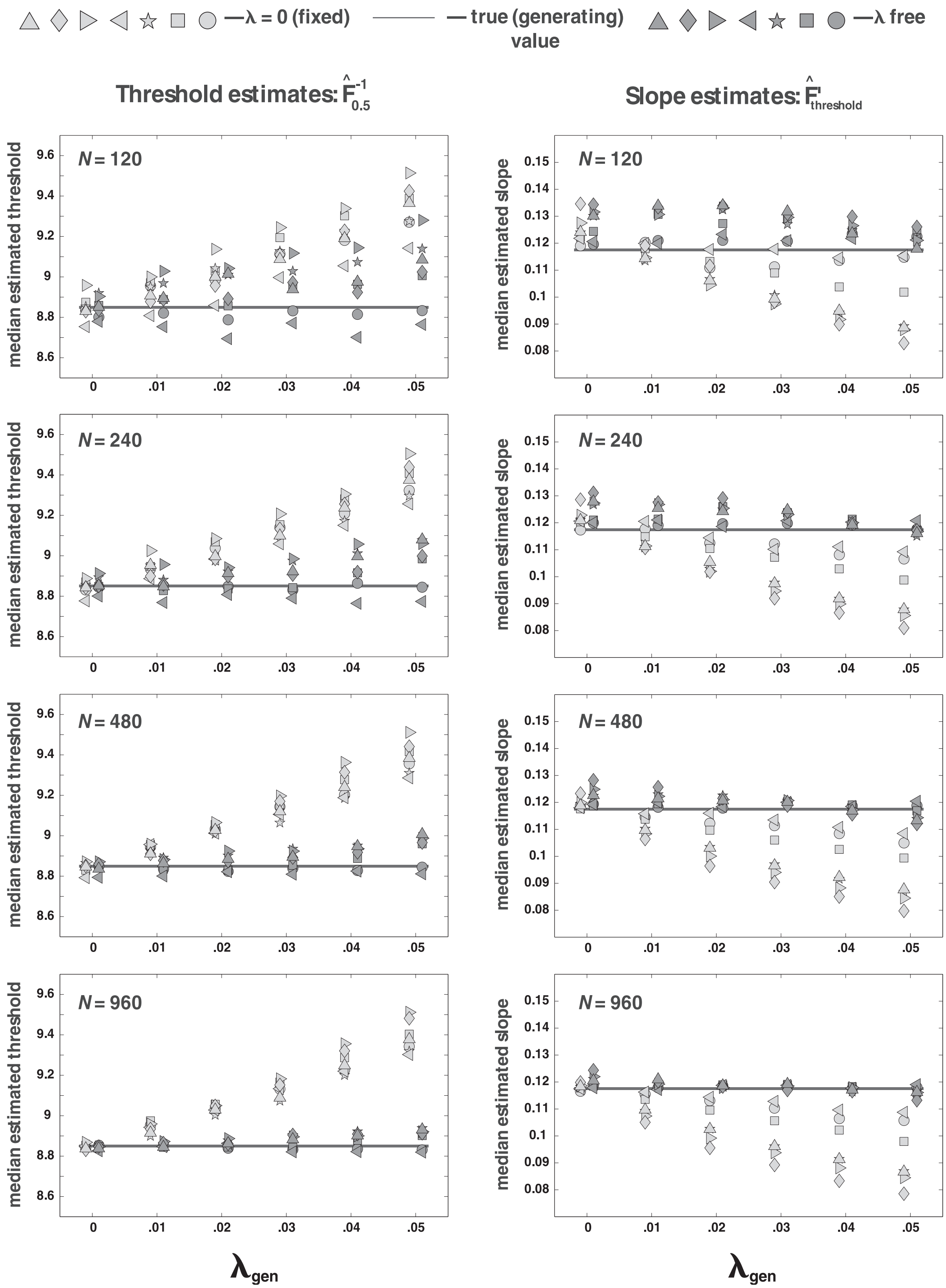

Figure 2. Results reported by Wichmann and Hill (2001). Median threshold and slope estimates across 2,000 simulations are shown for seven stimulus placement regimens (represented by the different symbol shapes) as a function of the generating lapse rate (λ_{gen}). All placement regimens contained six stimulus intensities with the total number of trials (N) evenly distributed among the different stimulus intensities. Light symbols correspond to fits in which the lapse rate was fixed at a value of zero, dark symbols correspond to fits in which the lapse rate was allowed to vary within a rectangular prior. (Reproduced from Figure 3 in: Wichmann and Hill [2001], with kind permission from Springer Science + Business Media B.V.).

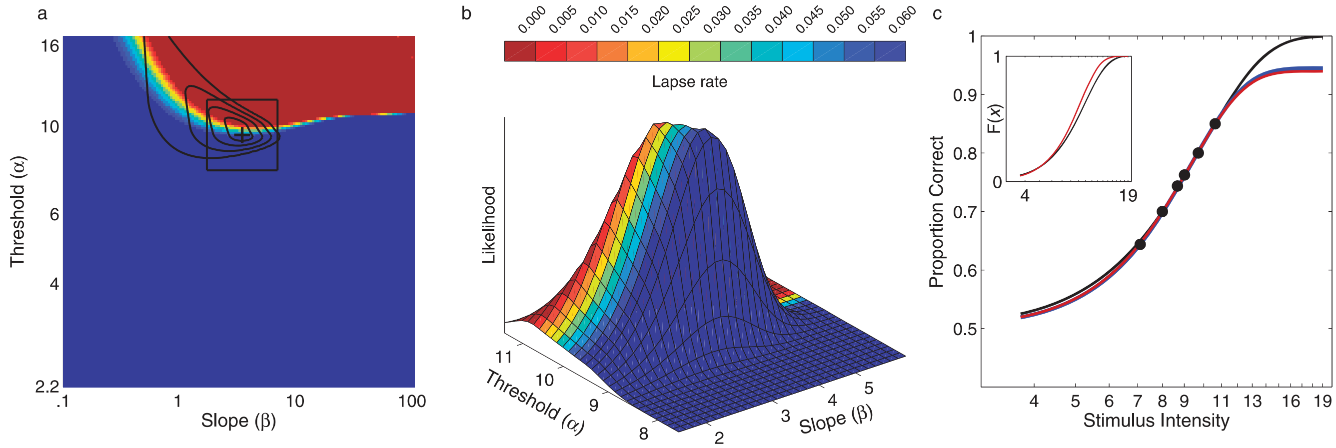

Figure 3. For each simulation, a brute-force search was first performed through a search grid containing 150 possible values for the threshold parameter, 150 possible values for the slope parameter and, in case the lapse rate was free to vary, 13 possible values for the lapse rate parameter. The full 3-D search grid thus contained 292,500 (150x150x13) PFs. In fits in which the lapse rate was fixed, the search grid was restricted to the (150x150) plane that corresponded to the fixed lapse rate. The best-fitting PF in the grid subsequently served as the seed for the iterative Nelder-Mead search. (A) Equal-likelihood contours for the example dataset shown in (C) across the threshold and slope values contained within the search grid. The color code here and in (B) indicates the value of the lapse rate of the PF with the highest likelihood. (B) Likelihood across threshold and slope for the region outlined by the square in (A). This region is centered on the PF in the search grid that has the highest likelihood. The grain of the grid as shown in the Figure corresponds to that of the search grid. (C) The generating curve with lapse rate equal to zero (black curve), some hypothetical data under placement regimen s1, the best-fitting PF contained in the search grid (blue curve, largely obscured by red curve), and the best-fitting PF resulting from the Nelder-Mead iterative search (red curve). Note that while the fitted curve corresponds closely (at least for the stimulus intensities included in s1) to the generating curve in terms of the probability of a correct response (ψ in Equation 1, function of α, β, γ, and λ), it does not in terms of F (inset; function only of α and β; see Equation 1). The code which accompanies this paper will produce a Figure such as this for any simulation performed in this paper as well as any of Wichmann and Hill's conditions not reported here.

a value within a narrow range, many lapse rate estimates will be equal to one of the limits of this range. Whether and to which degree this will be the case depends mainly on the number of observations, and the stimulus placement regimen (particularly whether the regimen includes placements at high stimulus intensities). Even when high stimulus intensities are included in the regimen, many lapse rate estimates may be equal to one of the limits of this range. Often the lapse rate estimate distribution is bimodal with peaks at both of the limits of the prior. Because I was interested in the distribution of the lapse rate estimates in Wichmann and Hill's (2001) simulations (Wichmann & Hill do not provide these) I attempted to replicate their results shown in Figure 2.

Method

Simulations

I repeated Wichmann and Hill's (2001) simulations that are shown in (this paper's) Figure 2 following their procedure closely (some details, such as the exact values for the stimulus placements were obtained from Hill, 2001). In order to avoid crowding of figures, results will be presented only for placement regimens s1, s6, and s7. These three placement regimens differ with respect to their placement of stimuli in characteristic ways that will prove to affect the behavior of parameter estimates systematically. Regimen s1 places stimuli exclusively near threshold level, regimen s6 is similar to s1 in that respect but also includes a single intensity at a performance level near asymptote. Regimen s7 has intensities around threshold level, one intensity at a very high performance level, but also includes intensities at intermediate levels. Matlab® code that can be used to perform all simulations and parameter estimations presented in this paper as well as the other placement regimens used by Wichmann and Hill is available here: www.palamedestoolbox.org/jovcode.html.

I also performed simulations in which stimulus placement was guided by the adaptive psi-method (Kontsevich & Tyler, 1999). The psi-method selects stimulus intensities on each trial such as to reduce uncertainty in the threshold as well as slope parameter estimates. Briefly, following each trial the psi-method derives a posterior probability distribution across (discrete) values for the threshold and slope parameters based on all previous trials and a user-provided prior

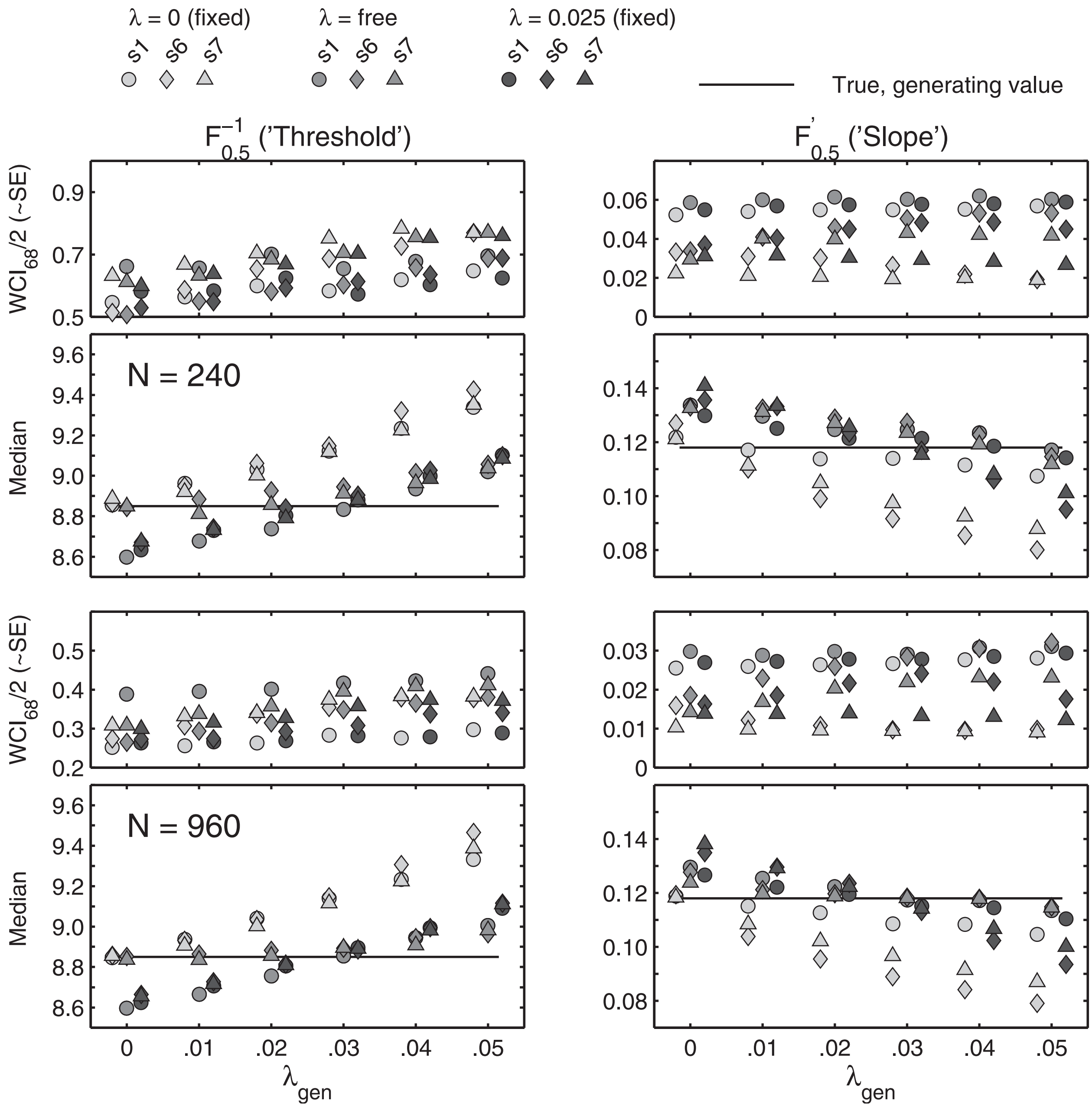

Figure 4. Attempted replication of the results shown in Figure 2. Also shown are 'half 68% confidence interval widths' (see text for details). Observed pattern of results for the simulations in which lapse rate was free to vary differs from that reported by Wichmann and Hill shown in their Figure 3 (reproduced here as Figure 2).

distribution. The stimulus intensity to be used on the next trial is then selected such that the expected entropy in the posterior distribution is minimized.

The range of values of the threshold parameter included in the psi-method's parameter space included 51 possible values spaced logarithmically between $F^{-1}_{0.1}$ ($=4.72$) and $F^{-1}_{0.9}$ ($=13.21$). The range of values of the slope parameter included in the psi-method's parameter space included 41 possible values spaced logarithmically between $\beta = 1$ ($F'_{0.5} = 0.050$ when $\alpha = 10$) and $\beta = 10$ ($F'_{0.5} = 0.360$ when $\alpha = 10$). A uniform prior across these parameter values was used. The range of possible stimulus intensities the psi-method could select from included 21 values spaced logarithmically between $F^{-1}_{0.1}$ ($=4.72$) and $F^{-1}_{0.999}$ ($=19.04$). The psi-method assumed a Weibull function, with lapse rate equal to 0.025 and a guess rate equal to 0.5. Note that the choice for assumed lapse rate and its correspondence to the generating value affect directly only the exact stimulus intensities used in the simulations, not the parameter estimates I report here as these are derived based on a maximum likelihood criterion in a

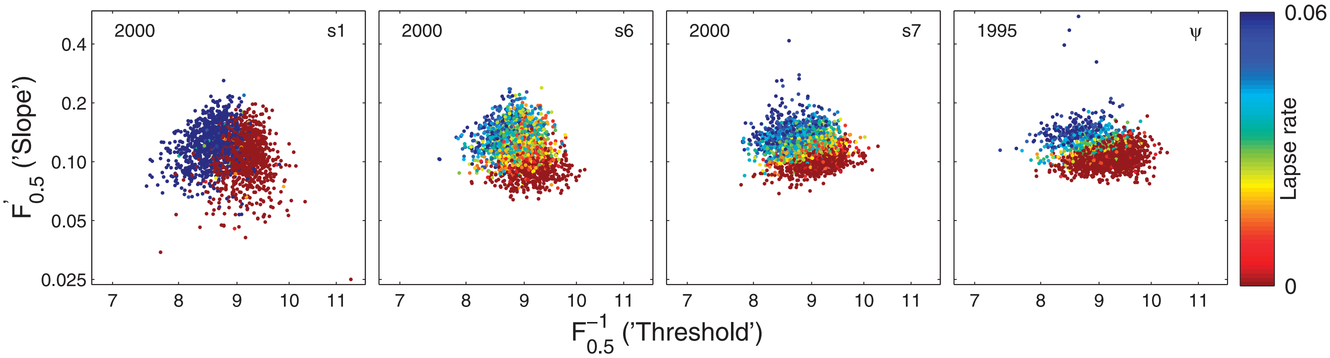

Figure 5. Scatterplots showing the relationship between parameter estimates for the three MOCS placement regimens as well as psi-method controlled placements (ψ). The generating lapse rate was 0.03 and N = 960. The number in each plot indicates the number of simulations (of 2000) resulting in parameter estimates contained within range of parameter values included in plots.

separate procedure. The psi-method was implemented using the Palamedes toolbox (Prins & Kingdom, 2009). In order to provide a general idea as to the placement of stimuli when the psi method is used, the stimulus placements combined across 10,000 simulations where the generating lapse rate equaled 0.03 and the number of trials was 960 is included in Figure 1.

The interdependencies among observations that are introduced by the nature of the psi-method (as well as any other adaptive method) introduce bias in parameter estimates in addition to any bias that may be introduced by other sources (see Kaernbach, 2001, for a detailed explanation of the mechanism behind this bias). In order to separate the effect of stimulus placement per se on the one hand and that of serial dependencies on the other, all psi-method trial runs were retested in a MOCS context (e.g., Kaernbach, 2001). That is, the exact series of stimulus intensities resulting from each psi-method run was used again to simulate a new set of responses. These sets of data would thus have identical placement to those resulting from the psi-method run but would not contain the serial dependency, which is inherently present in adaptive runs.

Parameter estimation

Maximum likelihood parameter estimates for each of the simulations were derived by the Palamedes toolbox (Prins & Kingdom, 2009). Details of the fitting procedure are described in Figure 3. Median threshold and slope estimates for Wichmann and Hill's (2001) MOCS placement regimens s1, s6, and s7 are presented in Figure 4 for N = 240 and N = 960. Parameter estimates are shown for fits in which the lapse rate was allowed to vary within the prior, the lapse rate was fixed at a value of zero and the lapse rate was fixed at a value of 0.025. In order to provide a measure of the variance of the parameter estimates 'half 68% confidence interval widths' ('$WCI_{68}/2$') are also shown in Figure 4. These are simply half the distance between the 16^{th} and 84^{th} percentile in the distribution of parameter estimates. Insofar as these distributions are normally distributed, these values are comparable to standard errors of estimate. Figure 5 shows scatterplots of parameter estimates obtained with a free lapse rate for s1, s6, s7, and psi-controlled placement regimens using a generating lapse rate of 0.03 and N = 960. Full distributions of parameter estimates in the form of histograms and scatterplots will be produced by the code that accompanies this paper for any of the simulations performed in this paper as well as any of the conditions in Wichmann & Hill's (2001) Figure 3 (reproduced here as Figure 2).

In Figure 4, my results for the conditions in which the lapse rate estimate was fixed at a value of zero are virtually identical to those of Wichmann and Hill (2001). With the lapse rate fixed at 0.025, I obtain results similar to those obtained by Wichmann and Hill in their Figure 5 (not reproduced here) where they fixed λ at values other than zero. Systematic and significant biases are observed in $F^{-1}_{0.5}$ as well as $F'_{0.5}$. The magnitude of bias depends primarily on the difference between the value of the generating lapse rate and the value assumed during the fit. In line with observations made by Klein (2001), fixing the lapse rate at a small (but greater than zero) value avoids the excessive biases in slope found when the lapse rate is fixed at a value of zero.

My results for the fits in which the lapse rate estimate was allowed to vary, however, are quite different from those reported by Wichmann and Hill. Whereas Wichmann and Hill's results indicate a lack of bias when the lapse rate estimate is allowed to vary within the prior window, my results instead do display a systematic bias. At low values of the generating lapse rate, threshold estimates are essentially unbiased as long as high values of F are included in the placement

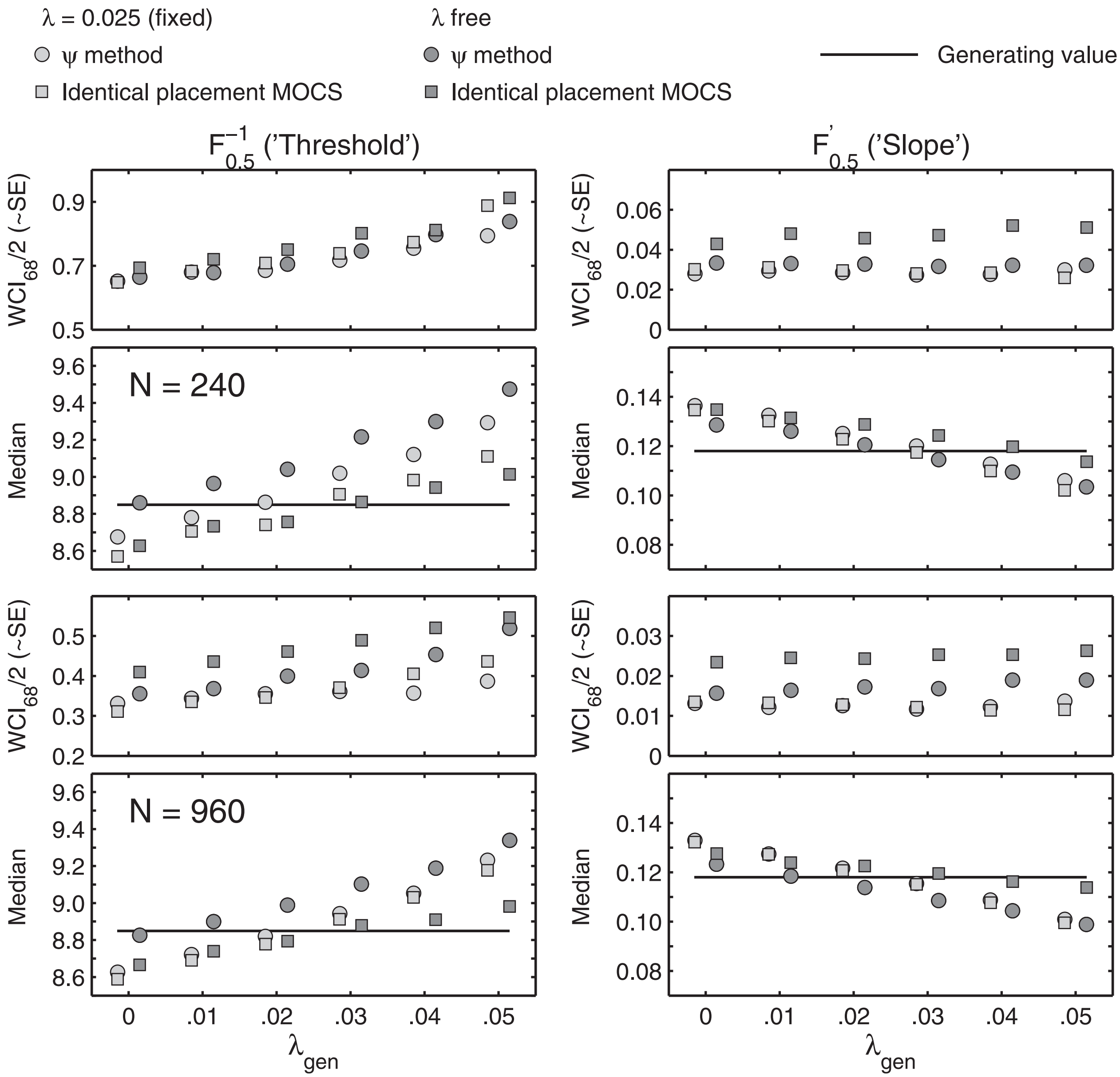

Figure 6. Median threshold and slope estimates when stimulus placement was guided by the adaptive psi-method (round symbols). Also shown are median parameter estimates derived from MOCS replications of psi-method controlled runs (square symbols). For each of the four sets of symbols in each graph are also shown 'half 68% confidence interval widths' (see text for details).

regimen (s6, s7 and not shown here: s3 and s5), but negatively biased when placement regimens are used, which do not include high values of F (s1 and, not shown here, s2 and s4). At high values of the generating lapse rate, *all* placement regimens (including those not shown here) lead to positively biased threshold estimates. Bias in slope estimates is small and consistent across placement regimens but varies systematically with generating lapse rate. From Figure 5 it is clear that the lapse rate estimate is correlated with both the threshold and slope parameters.

Results for the simulations in which stimulus placement was guided by the psi-method are presented in Figure 6 in a manner similar to Figure 4. When simulations using stimulus placements resulting from psi-method runs are repeated as MOCS (square symbols), the observed pattern of bias is comparable to that shown in Figure 4 for placement regimen s1 (which does not include high values of F; the psi-method runs also tend not to include stimuli placed at high stimulus intensities, see Figure 1). It is also clear that the use of the adaptive psi-method itself affects parameter estimate bias greatly (round symbols), especially when the lapse rate is free to vary and the number of trials is low. As noted above, this is due to the trial-to-trial interdependencies inherent in adaptive methods (Kaernbach, 2001).

Discussion

My results indicate that, contrary to Wichmann and Hill's (2001) claim, significant and systematic bias in parameter estimates remains when the lapse rate is allowed to vary during fitting. This bias is exacerbated when stimulus placement is governed by the adaptive psi-method. I will first discuss in some detail the source of this bias. I will then suggest and test an alternative estimation strategy, which suggests itself based on the considerations of the source of bias in parameter estimates.

Source of bias

I will focus my discussion on the threshold parameter estimates obtained using Wichmann and Hill's (2001) MOCS placement regimens. The argument I present in regard to the source of the bias in these threshold estimates applies equally well to the threshold estimates obtained with placements derived by psi-method but retested as MOCS. While the argument I present focuses on threshold parameter estimates it generalizes easily to slope estimates. I will also discuss the additional bias introduced when using the adaptive psi-method to control stimulus placement.

In order to understand the source of the bias observed when the lapse rate was free to vary, let us first note that two distinct patterns of bias in threshold are evident in Figure 3. For placement regimens s1 (and, not shown here, s2 and s4) bias in threshold is an approximately linear function of the generating lapse rate: Whereas at low generating lapse rates threshold estimates tend to underestimate the generating value, at high generating lapse rates threshold estimates tend to overestimate the generating value. Bias in threshold is near zero when the generating lapse rate equals 0.03. For placement regimens s6, s7 (and, not shown here, s3 and s5) on the other hand, threshold estimates are approximately unbiased at low generating lapse rates (up to about $\lambda_{\text{gen}} = 0.03$) but tend to overestimate the generating value when the lapse rate is high.

As will become evident from the argument I present below, the critical feature that sets s1 (as well as s2 and s4) apart from s6 and s7 (as well as s3 and s5) is that whereas the former do not include a stimulus placed at a high intensity, the latter do. The highest stimulus intensity used in regimen s1 was $F^{-1}_{0.7}$. Regimen s6 includes a stimulus placement at $F^{-1}_{0.99}$, s7 includes a stimulus placement at $F^{-1}_{0.998}$. Below I consider the sources of bias in some detail. I do so separately for the two observed patterns of bias.

Source of bias when placement regimen does not include a high intensity stimulus

Whereas this discussion focuses on bias observed under stimulus placement regimen s1, the argument generalizes to stimulus placements derived by the psi-method but retested as MOCS or any other regimen that does not include stimuli placed at a high intensity. Figure 3c displays the PF that was used as the generating function here and in Wichmann and Hill (2001) (i.e., $\alpha = 10$, $\beta = 3$ [corresponding to $F^{-1}_{0.5} = 8.85$ and $F'_{0.5} = .118$], $\gamma = 0.5$) with a lapse rate equal to 0 (black curve). A hypothetical data set is shown by the black symbols in the figure. When the lapse rate is allowed to vary within [0 0.06], the best-fitting PF is the red curve in the figure. Its estimate of the threshold parameter α equals 9.29, its estimate of the slope parameter β equals 3.38 and its estimate of the lapse rate parameter λ equals 0.06 (i.e., the upper limit on the lapse rate's prior). These values correspond to $F^{-1}_{0.5} = 8.33$ and $F'_{0.5} = 0.140$. As is clear from Figure 3c, despite having dissimilar parameter values, the generating curve and the best-fitting (red) curve are virtually identical *within the range covered by the s1 regimen* (which spans $F^{-1}_{0.3} = 7.09$ through $F^{-1}_{0.7} = 10.64$). It is important to note, however, that these curves are similar only in terms of probability of a positive (e.g., 'correct') response (i.e., ψ in Equation 1). The functions describing the underlying perceptual process (in which we are interested; F in Equation 1) are quite different, as shown in the Figure inset.

The red and black functions shown in Figure 3c in fact, merely the limits of an entire family of PFs that are virtually identical within the s1 placement range. These limits are defined by the boundaries placed on the lapse rate. In case an observer (real or simulated) generates data under placement regimen s1 and according to the generating PF with $\lambda = 0$, all PFs in the family bound by the two functions shown in Figure 3c are, within the tested range, virtually identical to the generating PF. Likewise, any dataset generated under the s1 placement regimen will have an entire family of PFs associated with it that will all have likelihoods very near the maximum in the likelihood function. Indeed, from Figure 3b, which shows the likelihood function across threshold and slope values for the hypothetical dataset shown in Figure 3c, it is clear that the likelihood function lacks a distinct peak but instead has a ridge corresponding to the family of PFs bound by the two functions shown in Figure 3c. Note that the ridge occurs because the lapse rate is free to vary (the value of the lapse rate of the PF with the highest likelihood is indicated by the color code). Note also

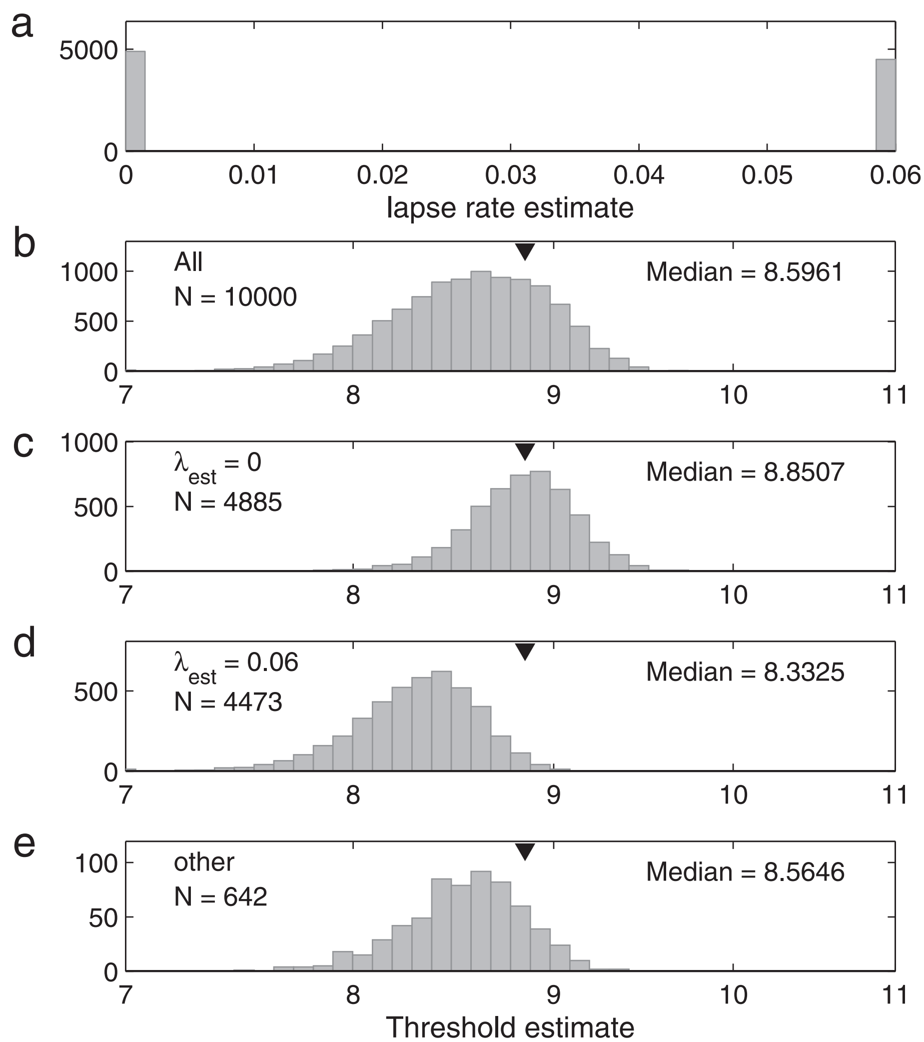

Figure 7. Results of 10,000 simulations in the s1 stimulus placement regimen when the lapse rate was free to vary. The generating lapse rate was zero. The generating threshold was $F_{0.5}^{-1} = 8.85$ and is indicated by the triangle in the Figures. (a) distribution of lapse rate estimates. (b) distribution of threshold estimates in all 10,000 simulations. (c) distribution of threshold estimates in those simulations in which $\hat{\lambda} = 0$. (d) distribution of threshold estimates in those simulations in which $\hat{\lambda} = 0.06$. (e) distribution of threshold estimates in the simulations in which $0 < \hat{\lambda} < 0.06$.

that the extent of the ridge is constrained by the limits placed on the lapse rate: The ridge would extend farther in both directions if the prior on the lapse rate would allow it.

It will be rare to find that the maximum in the likelihood function for data obtained under s1 occurs somewhere other than at either of the limits set by the prior on the lapse rate. Figure 7 shows parameter estimates for 10,000 simulations that were generated by the generating function with $\lambda = 0$ under placement regimen s1 with number of trials equal to N = 960. Figure 7a shows the distribution of lapse rate estimates for the 10,000 simulations. Nearly all lapse rate estimates (9,358 of the 10,000, or 93.6%) are at the limits of the prior, with about an equal number at each end of the prior (N = 4,885 at $\hat{\lambda} = 0$, N = 4,473 at $\hat{\lambda} =$ 0.06). Figure 7b shows a histogram of all 10,000 threshold estimates. From Figure 7b we note that thresholds are clearly biased (the generating threshold value is indicated in the figure by the triangle). The bias in threshold estimates is closely linked to the observed distribution of lapse rate estimates. Figure 7c shows the threshold estimates for the 4,885 simulations in which the lapse rate estimate equaled zero. For these simulations the lapse rate estimate was accurate (albeit mostly accidentally so, as I will argue) and we find that for this subset of simulations, the threshold estimates are unbiased. Effectively, these 4,885 simulations were fitted by a PF at the 'correct' limit of the family of PFs that would all fit these simulations about equally well.

However, Figure 7d shows the threshold estimates for the 4,473 simulations in which the lapse rate estimate equaled 0.06 (that is, those simulations for which the best-fitting PFs were at the other, 'incorrect' limit of the family of PFs). Note that the median threshold estimate for these 4,473 is very biased. The value of the median threshold estimate ($\hat{F}^{-1}_{0.5} = 8.33$) instead corresponds to the threshold of the PF which is the closest match to the generating PF within the range covered by the s1 regimen but which also has $\lambda = 0.06$ (cf. red curve in Figure 3c). The same pattern of results is observed in the median slope estimates. For the fits in which the lapse rate estimate was equal to zero, the median of slope estimates $\hat{F}'_{0.5}$ was equal to 0.118 (compare to the slope of the generating Weibull $F'_{0.5} = 0.118$). However, for the fits in which the lapse rate estimate was 0.06, the median of slope estimates $\hat{F}'_{0.5}$ was equal to 0.139 which corresponds closely to the slope of the red curve shown in Figure 3c ($\hat{F}'_{0.5} = 0.140$). Finally, in Figure 7e are shown the threshold estimates for the (relatively few) remaining simulations. The lapse rate estimates for these simulations are between those for the simulations in Figure 7c and 7d and so is the median threshold estimate for this subset of simulations.

When the 10,000 simulations were repeated but now with a generating lapse rate equal to 0.05, the distribution of lapse rates differed hardly from that shown in Figure 7a: 4,612 simulations resulted in a lapse rate estimate of 0 and 5,015 resulted in a lapse rate estimate of 0.06. The distribution of lapse rates estimates apparently has little to do with the generating lapse rate when placement regimen s1 is used. The median threshold for the 4,612 simulations resulting in a lapse rate estimate equal to 0 was $F^{-1}_{0.5} = 9.35$, that for the 5,015 simulations resulting in a lapse rate equal to 0.06 was $F^{-1}_{0.5} = 8.76$. These values correspond closely to the thresholds of the two functions which are the (prior-defined) limits of the family of PFs associated with the generating function with $\lambda = 0.05$. The leftmost scatterplot in Figure 5 shows the relationships among parameter estimates observed under placement regimen s1 in a different manner (the generating lapse rate in the figure equaled 0.03).

Source of bias when placement regimen includes a high intensity stimulus

While threshold estimates in stimulus regimens that include high stimulus intensities (s6, s7 and, not shown, s3 and s5) are essentially unbiased when the generating lapse rate is low, there is a systematic bias in these estimates when the generating lapse rate is high. It will prove worthwhile to consider the source of this asymmetry in some detail. When the placement regimen contains high stimulus intensities (i.e., those for which F [Equation 1b] is near unity) and the generating lapse rate equals zero, very few (if indeed any at all) incorrect responses will occur at the high stimulus intensity. Such a result would be consistent *only* with F having a value near unity *and* the lapse rate having a value near 0. As a result, when the generating lapse rate is zero and a placement regimen which includes a high stimulus intensity is used, lapse rate estimates will be at or near the ('correct') value of 0. Correspondingly, bias in threshold and slope will be minimal.

However, when the lapse rate is high, several incorrect responses are expected to be observed at the high intensity stimulus. This presents an ambiguous situation: The relatively high number of incorrect responses could be due *either* to a high lapse rate *or* to a low value of F (or some combination of these two factors). Stated more precisely, a relatively high number of incorrect responses at the high stimulus intensity will be consistent with a relatively broad family of functions, members of which will display a wide range of lapse rate values. The manner in which the three parameters trade off when placement regimen includes a high stimulus intensity is very apparent from Figure 5 (middle two panels: s6 and s7): High lapse rate estimates tend to go with threshold and slope estimates that combine to produce high perceptual performance (i.e., high F) at the high stimulus intensity (i.e., low threshold/high slope). Similarly, low lapse rate estimates tend to go with high threshold/shallow slope estimates. It might be noted in passing that, contrary to intuitive appeal perhaps, increasing the number of trials at the highest stimulus intensity will not do anything to resolve this ambiguity. That is, a proportion of, say, 5% incorrect responses at a high intensity will be consistent with either a low value of F or a high lapse rate regardless of the number of observations it is based on. The ambiguity must instead be resolved by obtaining accurate estimates of threshold and slope parameters through observations made at the lower stimulus intensities. Returning to our argument, the bias observed when the generating lapse rate is high arises because of the asymmetry of the window of allowed lapse rate estimates relative to a high generating lapse rate. Whereas the window allows those functions which have lapse rate values that are much *lower* than the generating value (which are coupled with upward biased threshold estimates), it does not allow those with lapse rate estimates that are (much) *higher* than the generating value. Overall, then, threshold estimates are biased upward when the generating lapse rate is high.

Bias introduced by using the adaptive psi-method

Let us first consider the bias in estimates when the stimulus placement was as in the psi-method but retested in an MOCS context and placement was thus not contingent upon previous responses. These results are indicated by the square symbols in Figure 6. The pattern of results for these conditions is very similar to that obtained in Figure 4 for placement regimen s1, which also did not include high stimulus intensities. That is, bias was obtained when the generating lapse rate was low or high and this bias was larger (especially when considering slope estimates) when the lapse rate was fixed at 0.025 compared to being free to vary. The source of bias in these conditions is the same as that discussed above for placement regimens that do not include high stimulus intensities. When we compare these results to those obtained in the original psi-method runs (circular symbols in Figure 6), however, it becomes clear that the contingencies have the effect of raising the median threshold estimates such that, while the median threshold estimates get closer to the generating value for some of the lower values of the generating lapse rates, the threshold estimates when considered overall become more biased. This effect is especially pronounced when N is low and the lapse rate is free to vary. As a matter of fact, the pattern of threshold estimates for these conditions is somewhat similar to that shown in Figures 2 and 4 for fits in which the lapse rate was fixed at a value of zero. Slope estimates, on the other hand, are overall less biased when the lapse rate is allowed to vary compared to being fixed.

Again, we find that bias in threshold estimates is closely linked with concurrent deviations of lapse rate estimates from the true, generating value. When the lapse rate is free to vary, lapse rate estimates mostly equal 0 (at $N = 960$, 73.4% of lapse rate estimates equal 0 when $\lambda_{\text{gen}} = 0$ and 52.6% of estimates equal 0 when $\lambda_{\text{gen}} = 0.05$). This finding seems to extend an observation made by Kaernbach (2001). Kaernbach demonstrated (and meticulously argued) that a bias in slope estimates results when an adaptive method is used that selects stimulus intensities such as to optimize measurement of the threshold but not that of the slope parameter. He further demonstrated that the bias in slope estimates is remedied when an adaptive method is used that selects stimulus intensities to optimize measurement of the slope as well as the threshold (see also Kontsevich & Tyler, 1999). The high degree of bias in lapse rate estimates (and the closely linked bias in threshold and slope parameter estimates) obtained here may thus be a result of the fact that the adaptive method selects stimulus intensities such as to optimize threshold and slope estimates, but not the lapse rate estimate. The general rule appears to be that unless an adaptive procedure optimizes stimulus selection for the estimation of a specific parameter, caution should be exercised when that parameter is subsequently estimated from the resulting observations.

Proposed alternative strategy

It has been suggested by some to include a relatively high proportion of trials at high intensities (e.g., Treutwein, 1995). Some of the placement regimens used in Wichmann and Hill (2001) and here did include high stimulus intensities and each of these presented 1/6 of the total trials at this high intensity. Whereas this indeed leads to essentially unbiased parameter estimates when the generating lapse rate is low, estimates are still biased when the generating lapse rate is high. I argued above that a critical element underlying this bias is that the source of incorrect responses at a high stimulus intensity is ambiguous. These incorrect responses may result from either a high lapse rate or a low value of F (or a combination of the two). As I have argued, merely increasing the number of observations at high intensities does nothing to resolve this ambiguity.

This observation suggests an alternative strategy to incorporating estimates of the lapse rate into our models. This strategy would involve including a proportion of trials at a stimulus intensity so high that it can be reasonably assumed that F at this intensity effectively equals unity. I will refer to such an intensity as an Asymptotic Performance Intensity or API. Critically, the model that is fitted would reflect the assumption that F equals unity at API. This strategy would remove the ambiguity as to the source of any incorrect responses at API. Otherwise, the fitting is performed as in the method advocated by Wichmann and Hill (2001). I call this strategy 'joint Asymptotic Performance Lapse Estimation' or jAPLE. In jAPLE, the model that is fitted is given by:

$$\begin{aligned} &\psi(x;\alpha,\beta,\gamma,\lambda) = 1-\lambda \quad \text{when } x = a \\ &\psi(x;\alpha,\beta,\gamma,\lambda) \\ &\qquad = \gamma + (1-\gamma-\lambda)\mathrm{F}(x;\alpha,\beta) \quad \text{otherwise.} \end{aligned} \tag{2}$$

In Equation 2, stimulus intensity a is an API. In effect, errors made at $x = a$ will be unambiguously attributed to lapses. Note that under jAPLE, observations made at intensities other than $x = a$ also contribute to the estimation of the lapse rate.

A second alternative method of fitting datasets in which a proportion of trials is collected at an API is a two-step procedure. In the first step, the lapse rate estimate is derived based solely on observations made at the API, under the assumption that incorrect responses observed there are exclusively due to lapses. The maximum-likelihood estimate of the lapse rate thus would simply correspond to the proportion of incorrect responses observed at the API. In the second step, the threshold and slope are estimated from observations made at the non-API intensities while fixing the lapse rate at the value obtained in the first step. I will refer to this second strategy as isolated-Asymptotic Performance Lapse Estimation or 'iAPLE.' Note that under iAPLE, observations made at intensities other than API do not contribute to the estimate of the lapse rate.

In order to provide a brief demonstration and test of the proposed methods, I modified Wichmann and Hill's (2001) placement regimens simply by changing the highest stimulus intensity included in each of the regimens (whatever that intensity was) to $F^{-1}_{0.9999}$ (that is, to a stimulus intensity at which an incorrect response would almost exclusively be due to a lapse).

I also tested the proposed methods with data collected using the adaptive psi-method. For the purposes of comparison I used the same number of total trials (i.e., $N = 240$ and 960) in these simulations as I did above. However, only 5/6 of the total of N trials were collected using the adaptive psi-method, the remaining 1/6 were simulated using a constant stimulus intensity of $F^{-1}_{0.9999}$. Here too, all simulations were repeated as MOCS in order to isolate the effect of intertrial stimulus-response interdependencies. Note that the total number of trials used in all simulations remains identical to that used above.

Parameter estimates for all simulations were then derived under the jAPLE and the iAPLE methods. As a control, parameter estimates were also derived using the method proposed by Wichmann and Hill. In the remainder, I will refer to Wichmann and Hill's proposed method as nAPLE ('non-Asymptotic Performance Lapse Estimation'). Results for modified placement regimens s1, s6, and s7 are presented in Figure 8 alongside results using the original placement regimens fitted nAPLE (which were also shown in Figure 4). Results for the psi-method controlled stimulus placements are presented in Figure 9. Figure 10 shows scatterplots for the modified placement regimens and nAPLE, jAPLE, and iAPLE fitting schemes (generating lapse rate $= 0.03$ and $N = 960$).

Results indicate that, in terms of bias, the jAPLE and iAPLE methods outperform nAPLE at both values of N tested here. As a matter of fact, only at $N = 240$ does either method seem to display a mild but systematic bias which is dependent on placement regimen. While Wichmann and Hill's method of estimating parameters using the modified placement regimens outperforms their method using the original placement regimens, bias (albeit small) remains even at the highest value of N tested. In terms of the precision of estimates, however, performance suffers under the modified placement regimens in some conditions. This is especially evident for slope estimates under placement regimen s1 when jAPLE and iAPLE are used. Since jAPLE and iAPLE assume that incorrect responses at the API result only from lapses, observations taken at API do not contribute directly to the estimates of threshold or slope. Especially under s1, the remaining five intensities used to estimate threshold and slope are poorly placed to achieve high precision in the estimate of the slope.

Inspection of Figure 10 indicates that a trade-off among the three parameters is much more apparent under the nAPLE fitting scheme compared to the methods proposed here, especially under (modified) sampling schemes s1 and s6. As discussed here, incorrect responses made at a high stimulus intensity may be due either to lapses or to genuine perceptual misses (i.e., low value of F). In order to disambiguate the source of such errors, we must obtain accurate estimates of both threshold and slope from observations made at lower intensities. Modified placement regimens s1 and s6 cannot provide these: The five intensities that are not at API are all near threshold value in both s1 and s6.

A very small, but apparently systematic bias remains, even at high N, when the results from the psi-method runs are fitted directly. This bias is much more pronounced when the nAPLE method is used compared to the methods proposed here. At high N, parameter estimates are essentially unbiased when intertrial dependencies are removed by retesting the psi-method stimulus placements simulations using MOCS.

A few concluding remarks

Wichmann and Hill (2001) chose to report their results in terms of $F^{-1}_{0.5}$ and $F'_{0.5}$, rather than α and β (see Equation 1). The metrics used by Wichmann and Hill have the advantage of allowing numerical comparison of parameter values across different forms of F (Weibull, Logistic, etc.). However, $F^{-1}_{0.5}$ and $F'_{0.5}$ are both non-linear functions of both α and β, and it is the values of α and β which are estimated in the maximum-likelihood estimation procedure. Thus, while maximum-likelihood estimators have the desirable property of being asymptotically unbiased (e.g., Edwards, 1972), this property would apply only to α and β, not to $F^{-1}_{0.5}$ and $F'_{0.5}$. Moreover, since $F^{-1}_{0.5}$ and $F'_{0.5}$ are both functions of both α and β, any bias in *either* α *or* β would result in bias for *both* of $F^{-1}_{0.5}$ and $F'_{0.5}$. Since my results directly challenge the integrity of Wichmann and Hill's results I have chosen to report my results in terms of $F^{-1}_{0.5}$ and $F'_{0.5}$ also. However, my pattern of results would be the same whether expressed in terms of $F^{-1}_{0.5}$ and $F'_{0.5}$ or in terms of α and β. That is, like $F^{-1}_{0.5}$ and $F'_{0.5}$, α and β are both biased when estimated by the method proposed by Wichmann and Hill and, like $F^{-1}_{0.5}$ and $F'_{0.5}$, α and β are both not (noticeably) biased when the methods I propose are used and the number of observations is sufficient.

Based on the results of my simulations one should consider a number of issues before deciding whether to

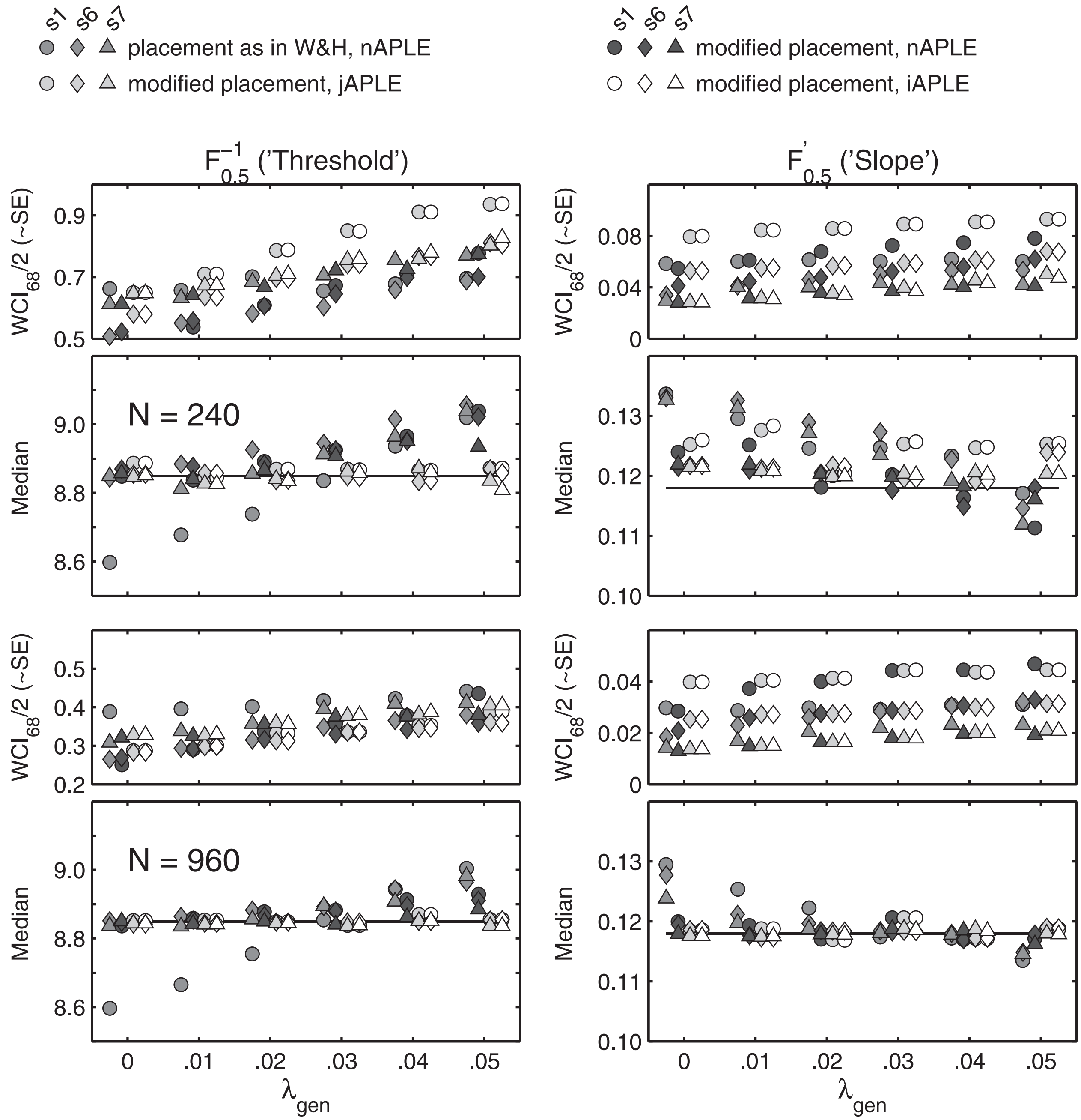

Figure 8. Median parameter estimates obtained using the methods proposed here (jAPLE and iAPLE). Modified placement regimens were as shown in Figure 1 except that the highest stimulus intensity was changed to $F^{-1}_{0.9999}$. Also shown are median parameter estimates derived from the same simulations using the method advocated by Wichmann and Hill (nAPLE), as well as median parameter estimates using the original placement regimens fitted nAPLE (i.e., these are identical to those shown in Figure 4 and replicated here only for the purpose of easy comparison). Also shown for each condition are 'half 68% confidence intervals widths' (see text for details).

allow the lapse rate to vary while fitting the PF. First, it is important to realize that bias in threshold and slope remains even when we allow the lapse rate to vary. With the lapse rate allowed to vary, the bias in slope estimates is largely independent of the sampling scheme chosen, while bias in threshold estimates is very dependent on sampling scheme (at least when the generating lapse rate is small). The opposite is true of biases when the lapse rate is fixed at a small (but greater than zero) value. In that case threshold bias is largely independent of sampling scheme while the bias in slope does depend on sampling scheme. Second, the degree of bias is very much dependent on the specific limits set on the prior window. Just as it is no coincidence that threshold and slope estimates were unbiased when the generating value of the lapse rate equaled its fixed,

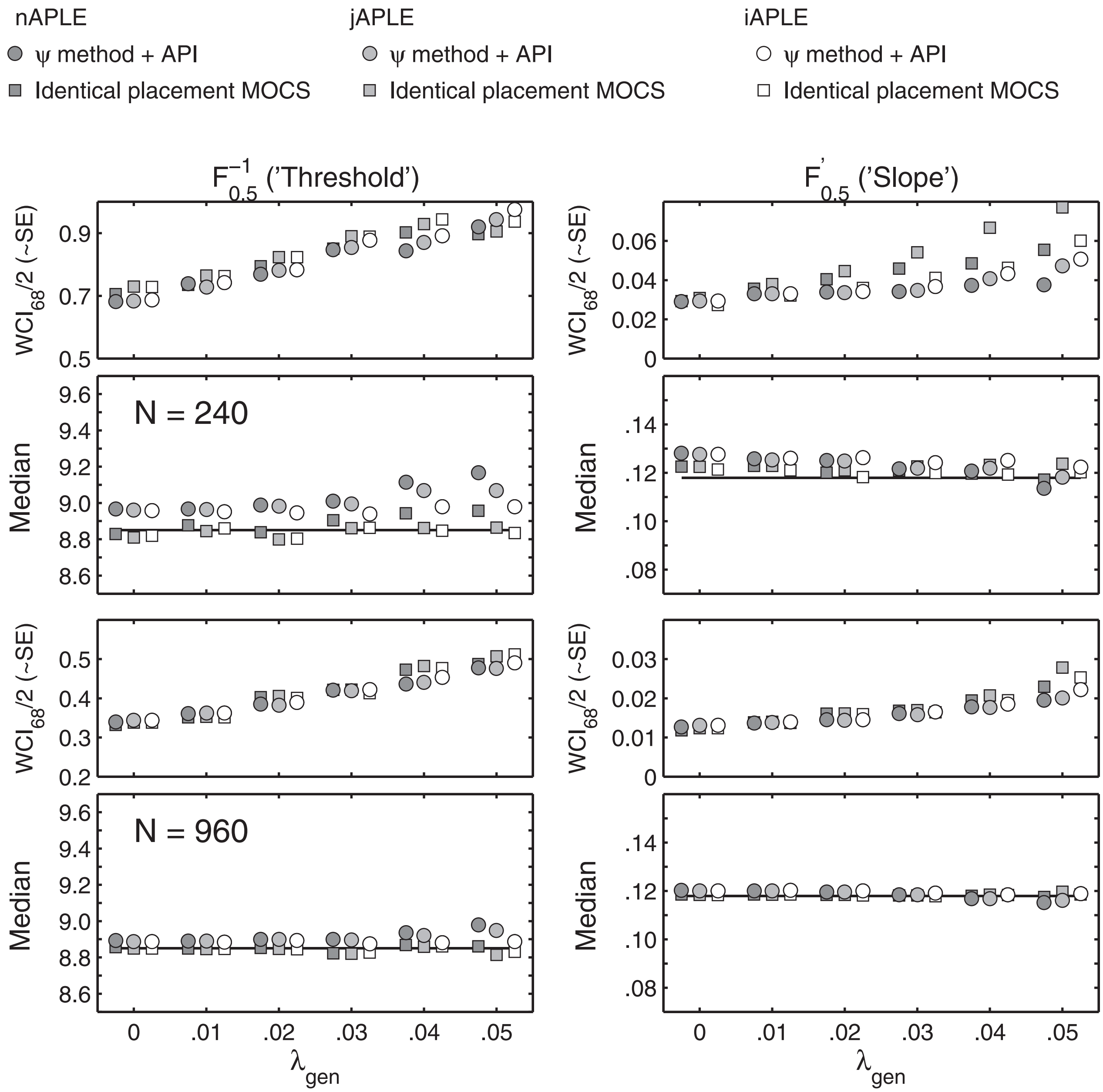

Figure 9. Median parameter estimates obtained using the methods proposed here (jAPLE and iAPLE). Stimulus placement of 5/6 of the trials was governed by the psi-method. The remaining trials used stimulus placement at $F^{-1}_{0.9999}$. Also shown are median parameter estimates derived from the same simulated datasets using the method advocated by Wichmann and Hill (nAPLE). The results obtained from data obtained from the psi-method runs directly are shown as round symbols, while square symbols display results from simulations in which stimulus placements obtained in each psi-method run were retested as MOCS. Also shown for each condition are 'half 68% confidence interval widths' (see text for details).

assumed value, it is also no coincidence that none of the sampling schemes produce a significant bias when the generating lapse rate equaled 0.03 (i.e., midway between the limits of the prior). Third, when an adaptive method is used, bias in threshold may actually be much greater when the lapse rate is allowed to vary compared to being fixed at a small, non-zero value. Fourth, when we allow the lapse rate to vary a stimulus at an API intensity (that is, one at which performance has reached asymptotic level) should be included in the sampling scheme even when, otherwise, stimulus intensities are selected by an adaptive method. The placement regimen should contain additional stimuli placed such as to obtain accurate estimates of both threshold *and* slope. Fifth, when an API stimulus is included one should consider using the proposed jAPLE or iAPLE fitting method. It is important to realize, however, that these methods may not always be

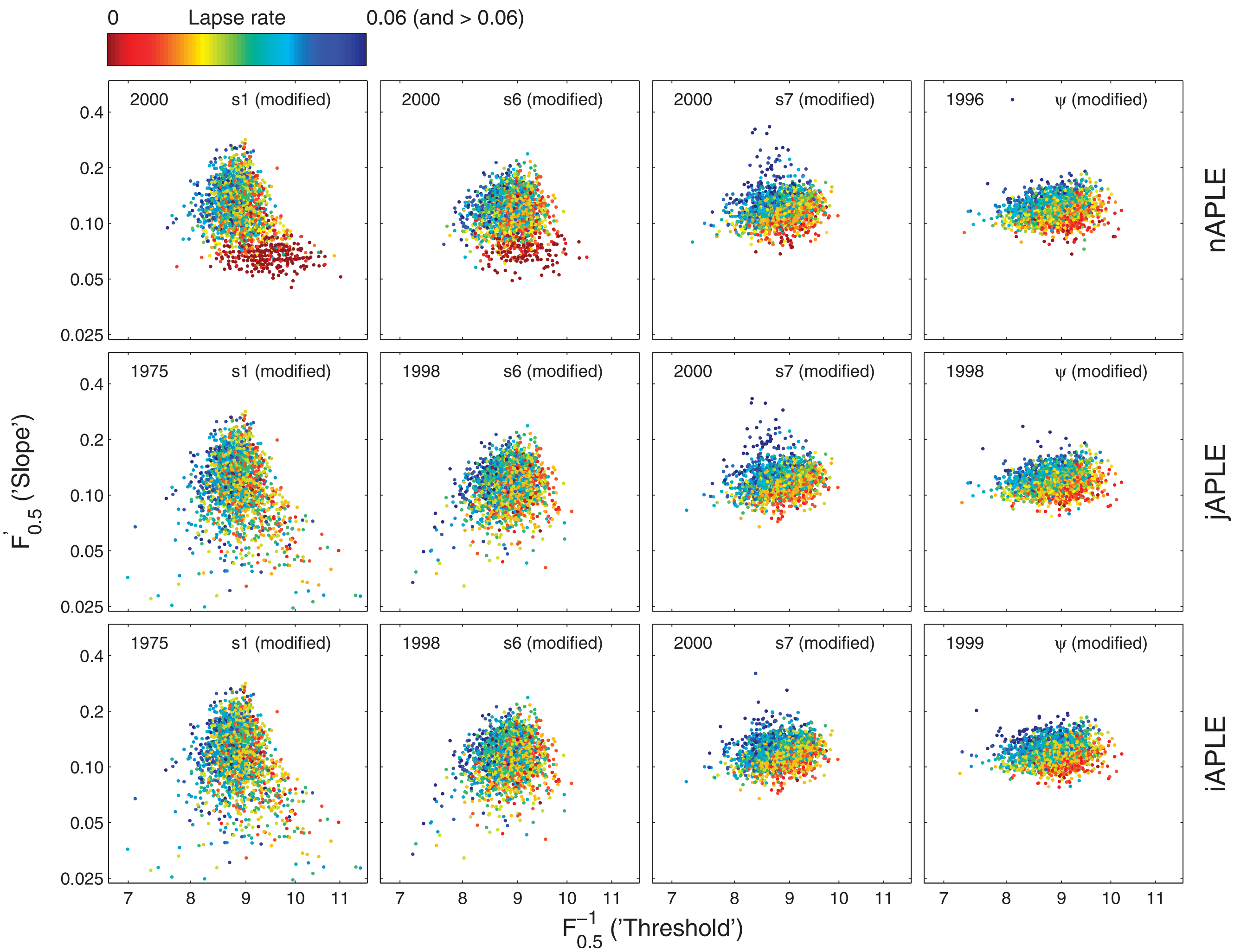

Figure 10. Scatterplots showing the relationship between parameter estimates for the three modified MOCS placement regimens as well as modified psi-method controlled placements using the different fitting schemes. The generating lapse rate was 0.03 and N = 960. Number in each plot indicates the number of simulations (of 2,000) resulting in parameter estimates contained within range of parameter values included in plots.

possible to implement since the maximum achievable stimulus intensity may not be at asymptotic performance. Great care should be taken to ensure that performance has indeed reached an asymptotic level at the stimulus intensity chosen as API.

One other possible strategy of dealing with the issue of the lapse rate might prove valuable. In most practical cases, bias in the threshold and slope estimates per se is of no concern. Of theoretical concern, generally, is not the absolute value of a threshold or slope, but rather whether differences in parameter values exist between experimental conditions. In such cases, rather than fitting thresholds and slopes to different experimental conditions individually, we may reparametrize our thresholds and slopes (e.g., Yssaad-Fesselier & Knoblauch, 2006; Kingdom & Prins, 2010). For example, in a two-condition experiment, we may reparametrize our thresholds into a parameter corresponding to the sum of the thresholds in the two conditions and a parameter corresponding to the difference between thresholds. Of theoretical concern in most research will be the value of the 'difference parameter' while that of the 'sum parameter' will, generally, have few theoretical implications. Of interest, then, is whether the difference parameter is subject to bias when assumptions regarding the lapse rate are violated. The issue of bias in the sum parameter and difference parameter is the focus of current research in my lab.

Acknowledgments

The author is grateful to Stan Klein, Felix Wichmann, Mike Landy, Miguel García-Pérez, and an anonymous reviewer for valuable comments on earlier versions of this paper. The author thanks Felix Wichmann also for his gracious permission to reproduce Figure 3 of Wichmann and Hill (2001).

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

Footnote

[1]Wichmann and Hill (2001) refer to the interval of allowed lapse rates as a 'Bayesian prior.' Constraining the lapse rate estimates to an interval that reflects the subjective belief concerning the likely values of the lapse rate does indeed embody a critical feature of Bayesian reasoning. However, the estimation of parameter values in Wichmann and Hill and here is performed by (constrained) maximum-likelihood estimation, not by Bayesian estimation. For that reason I will refer to the interval of allowed lapse rates simply as the 'prior window' or the 'prior.'

References

Edwards, A. F. W. (1972). *Likelihood*. Baltimore, MD: Johns Hopkins University Press.

García-Pérez, M. A., & Alcalá-Quintana, R. (2005). Sampling plans for fitting the psychometric function. *The Spanish Journal of Psychology, 8*(2), 256–289.

García-Pérez, M. A., & Alcalá-Quintana, R. (2007). The transducer model for contrast detection and discrimination: Formal relations, implications, and an empirical test. *Spatial Vision, 20*(1–2), 5–43.

Hill, N. J. (2001). Testing hypotheses about psychometric functions–an investigation of some confidence interval methods, their validity, and their use in the evaluation of optimal sampling strategies. Doctoral thesis, University of Oxford, Oxford, UK. Internet site: http://bootstrap-software.org/psignifit/publications/hill2001.pdf (Accessed May 21, 2012).

Kaernbach, C. (2001). Slope bias of psychometric functions derived from adaptive data. *Perception & Psychophysics, 63*(8), 1389–1398.

Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A practical introduction*. London: Academic Press: An imprint of Elsevier.

Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception and Psychophysics, 63*(8), 1421–1455.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research, 39*(16), 2729–2737.

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. [PubMed].

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

Swets, J. A. (1961). Is there a sensory threshold? *Science, 134*, 168–177.

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

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

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

Yssaad-Fesselier, R., & Knoblauch, K. (2006). Modeling psychometric functions in R. *Behavior Research Methods, 38*(1), 28–41.