Articles in PresS. J Neurophysiol (January 13, 2016). doi:10.1152/jn.00318.2015

A Quantitative Confidence Signal Detection Model: 1. Fitting Psychometric Functions

Short Title: Fitting Psychometric Functions via a Confidence Model

Yongwoo Yi[1,2]
Daniel M. Merfeld[1,2]

[1] Jenks Vestibular Physiology Lab
Massachusetts Eye and Ear Infirmary
Boston, MA, USA

[2] Department of Otolaryngology
Harvard Medical School
Boston, MA, USA

Address for correspondence:
Daniel M. Merfeld
Jenks Vestibular Physiology Lab
Room 421, Mass. Eye and Ear Infirmary
Boston, MA, USA 02114
dan_merfeld@meei.harvard.edu
fax +1 (617) 573-5596
phone +1 (617) 573-5595

Address for reprint requests and other correspondence: D. Merfeld, (e-mail: dan_merfeld@meei.harvard.edu). Interested parties are also invited to contact DM to request software used to fit the data.

Abstract

Perceptual thresholds are commonly assayed in the lab and clinic. When precision and accuracy are required, thresholds are quantified by fitting a psychometric function to forced-choice data. The primary shortcoming of this approach is that it typically requires 100 trials or more to yield accurate (i.e., small bias) and precise (i.e., small variance) psychometric parameter estimates. We show that confidence probability judgments combined with a model of confidence can yield psychometric parameter estimates that are markedly more precise and/or markedly more efficient than conventional methods. Specifically, both human data and simulations show that including confidence probability judgments for just 20 trials can yield psychometric parameter estimates that match the precision of those obtained from 100 trials using conventional analyses. Such an efficiency advantage would be especially beneficial for tasks (e.g., taste, smell, and vestibular assays) that require more than a few seconds for each trial, but this potential benefit could accrue for many other tasks.

Keywords – thresholds, decision-making, confidence rating, confidence calibration,

Introduction

Measuring thresholds is probably the most common psychophysical procedure in use today; applications range from experimental psychology to neuroscience to economics to engineering. Fitting psychometric functions using categorical data analyses (Agresti 1996) that describe the relationship between a stimulus characteristic (e.g., amplitude) and a subject's forced-choice categorical responses provides a standard approach used to estimate thresholds (Green and Swets 1966; Macmillan and Creelman 2005).

A recent comprehensive analysis (Garcia-Perez and Alcala-Quintana 2005) concluded that only maximum likelihood methods should be used when accuracy and precision of psychometric function fit parameters is important and, further, showed that more than 100 forced-choice trials are generally required to yield acceptable fit parameter estimates. Because such perceptual threshold tests are common and because many trials are needed to yield accurate and precise psychometric fits, studies spanning 50 years (Garcia-Perez and Alcala-Quintana 2005; Green 1990; Hall 1968; Hall 1981; Harvey 1986; Kaernbach 1991; Kontsevich and Tyler 1999; Leek 2001; Lim and Merfeld 2012; Pentland 1980; Shen et al. 2015; Shen and Richards 2012; Taylor and Creelman 1967; Treutwein 1995; Watson and Pelli 1983; Watt and Andrews 1981; Wetherill and Levitt 1965; Wichmann and Hill 2001a; b) have reported efforts to improve threshold test efficiency (i.e., to reduce the number of trials), but only modest efficiency improvements have accumulated. (For a brief review of these papers, see Karmali et al. (Karmali et al. 2015), which also presents a theoretic analysis of these sampling issues – accompanied by both simulations and human data.) This is probably due to the fact that binary/binomial distributions inherently have high variability at near-threshold stimulus levels – where the maximal information can be attained on each trial (Wetherill 1963).

While forced choice procedures are simple and robust; many subjects also know how confident they are for each response. Confidence is a belief in the validity of what we believe and is widely considered a form of metacognition (Drugowitsch et al. 2014; Fleming and Dolan 2012; Grimaldi et al. 2015; Lau and Rosenthal 2011) because it involves self-monitoring of perceptual performance. In other words, confidence reflects self-assessment of the conviction in a decision.

Confidence has been studied in humans using a variety of techniques (Balakrishnan 1999; García-Pérez and Alcalá-Quintana 2011; 2012; Garcia-Perez and Peli 2014; Hsu and Doble 2015; Okamoto 2012; Sawides et al. 2013) including probability judgments (Baranski and Petrusic 1994; Björkman 1994; Ferrell 1995; Ferrell and McGoey 1980; Juslin et al. 1998; Keren 1991; Lichtenstein et al. 1982; Stankov 1998; Stankov et al. 2012; Suantak et al. 1996). In fact, as noted in a recent review (Grimaldi et al. 2015), confidence probability judgments (i.e., confidence ratings provided using a nearly continuous scale between 0 and 100% or 50% and 100%) provide the most common assessment of confidence.

One common use of confidence recordings is in “confidence calibration” studies where confidence is compared to actual performance, where a data set may be classified as “well-calibrated” or classified as indicative of “overconfidence” or “underconfidence” (Baranski and Petrusic 1994; Björkman 1994; Ferrell and McGoey 1980; Juslin et al. 1998; Keren 1991; Lichtenstein et al. 1982; Stankov 1998; Suantak et al. 1996). Specifically, imagine that a subject reported 90% confidence that a given motion was rightward for 10 separate trials at a given stimulus level. On average, perfect calibration (Bjorkman et al. 1993; Ferrell 1995; Keren 1991; Lichtenstein et al. 1982; Stankov et al. 2012) of these confidence reports is assumed when 9 out of 10 of these trials are in the rightward direction, while overconfidence would be indicated by 5 out of 10 being rightward.

To our knowledge, probability judgments have never before been directly used to help estimate psychometric function parameters. Typically, confidence is not recorded. Sometimes a confidence rating (e.g., "uncertain") is recorded (Balakrishnan 1999; García-Pérez and Alcalá-Quintana 2011; 2012; Garcia-Perez and Peli 2014; Hsu and Doble 2015; Okamoto 2012; Sawides et al. 2013) and used as part of a psychometric fit procedure, but these approaches do not model how confidence quantitatively changes as the stimulus is varied. Instead these approaches include one additional decision boundary for each added category (e.g., "uncertain") – and typically add one free parameter to the fit algorithm for each additional decision boundary.

We now introduce our confidence signal detection (CSD, pronounced "kissed") model, which combines a confidence function (Fig. 1D) with a standard signal detection model (Fig. 1A-C). We use an example to illustrate the relationship between psychometric functions and confidence for a direction-recognition forced-choice task. A typical perceptual direction-recognition paradigm begins with well-controlled stimuli that are either positive or negative; the subject's task is to determine whether the motion is positive ("rightward") or negative ("leftward"). Typically, the stimuli provided to a subject (Fig. 1A) are well controlled (i.e., have little variation). The standard signal detection model suggests that neural noise contributes to perception (Goris et al. 2014; Tolhurst et al. 1981), which is represented by the probability density function (PDF) shown in Fig. 1B. Signal detection theory advocates that a single sample from this probability distribution - often called the decision variable – is available to the subject for each trial. If the decision variable sampled from this PDF for an individual trial is negative, the subject reports negative (e.g., leftward) motion and if the sampled decision variable is positive, the subject reports positive (e.g., rightward) motion. For the stimulus and noise PDF shown, positive motion will, on average, be reported 84% of the time. When this process is repeated for different stimulus amplitudes, it leads to a fitted psychometric function, ($\hat{\Psi}(x)$), that represents subject performance as a function of stimulus amplitude (Fig. 1C). Large positive stimuli will almost always be correctly reported as positive, and large negative stimuli will almost always be correctly reported as negative. Stimuli in between lead to the sigmoidal shape shown. With enough data, this fitted psychometric function ($\hat{\Psi}(x)$) is assumed to converge to a psychometric function that is representative of the subject's underlying noise distribution, $\hat{\Psi}(x) \approx \Psi(x)$.

So far, we have simply introduced the conventional signal detection approach; we now add a confidence model. If we happen to sample a very positive decision variable for one trial (e.g., if the stimuli were very large), we should have high confidence that the motion was positive. If we happen to sample a positive decision variable near the decision boundary on another trial, we should again decide that the motion was positive but have much less confidence in that decision. Like the empiric relationship captured by a psychometric function ($\hat{\Psi}(x)$), a quantitative empiric relationship between confidence and the stimulus can be represented by a confidence function ($\hat{\chi}(x)$). As for the psychometric function, with enough data, this empiric confidence function is assumed to be representative of neural processes that can be captured by a confidence function, $\hat{\chi}(x) \approx \chi(x)$. Figure 1D shows three example confidence functions that are each modeled as Gaussian cumulative distribution functions (CDFs); the solid curve represents well-calibrated confidence ($\chi(x) = \Psi(x)$), the dashed curve represents over-confidence, and the dotted curve represents under-confidence.

As will be described in detail in the Methods section entitled “Confidence Maximum Likelihood Fit Technique”, we utilize this CSD model to help improve psychometric parameter estimates. More specifically, we present a new confidence analysis technique that utilizes this CSD model. We will introduce, develop, and investigate this model using previously published analytic, simulation, and experimental approaches. To help evaluate the contributions that confidence can make to psychometric function estimation, we report: (1) human studies for a direction-recognition task in which the subjects were required to report whether they rotated toward their left or right, and (2) simulation results for psychometric functions that range from 0 to 1, which are used for direction-recognition data analysis (Chaudhuri et al. 2013; Merfeld 2011). We report that psychometric functions estimated using confidence probability judgments require about 5 times fewer trials to yield the same performance as conventional forced-choice psychometric methods.

Such improved test efficiency might be realized for any forced-choice task where confidence can be reported but may be especially important for perceptual tasks involving olfaction, gustation, or equilibrium (or any other task where individual trials take tens of seconds) (Linschoten et al. 2001) as well as for clinical applications where more efficient and/or more precise perceptual measures could lead to improved patient diagnoses (Merfeld et al. 2010). Nonetheless, while the promise of a 5-fold improvement in efficiency is appealing, we do not, at this time, advocate replacing standard forced-choice analyses with this confidence probability-judgment analyses because confidence in perceptual decisions is not well understood and our CSD model is not validated. We instead present this CSD model and new analysis method as deserving of further study.

Methods

Confidence Maximum Likelihood Fit Analysis

This section presents the new maximum likelihood analysis developed to help estimate psychometric function fit parameters. This technique simultaneously fits both a conventional psychometric function ($\hat{\Psi}(x)$) and a confidence function ($\hat{\chi}(x)$) to confidence probability judgments. Appendix A presents flow charts that outline this fitting technique. The specific model we use is presented via the flow chart on the left side of appendix A and a generalized flow chart is provided on the right side.

To introduce the new technique we assume that the fitted psychometric function can be represented by a Gaussian cumulative distribution function (ϕ) having two fit parameters ($\hat{\mu}, \hat{\sigma}$):

$$\hat{\Psi}(x) = \phi(x; \hat{\mu}, \hat{\sigma}) \tag{1}$$

where $\hat{\mu}$ represents shifts in the psychometric function (i.e., mean value of the noise distribution) and $\hat{\sigma}$ represents the width of the psychometric function (i.e., standard deviation of the noise distribution), which is often referred to as the threshold for direction-recognition tasks. Assuming that subjects based their confidence assessment on the signal used to make their decision, we modeled the fitted confidence function as a Gaussian cumulative distribution function having one additional free parameter, a confidence-scaling factor ($\hat{k}$) that scales this average confidence function to account for under-confidence or over-confidence, as previously demonstrated in Fig. 1D:

$$\hat{\chi}(x) = \phi(x, \hat{\mu}, \hat{k}\hat{\sigma}) \qquad (2)$$

We assume a Gaussian confidence function for simplicity, but other shapes of the confidence function were investigated via simulations to evaluate the impact of this assumption. Noise was not explicitly included in this relationship; noise may be present in the mapping from a decision variable to the confidence response, and we evaluate the impact of additive noise via simulations. While the rest of this paper focuses on using this CSD model to fit data, Figure 2 helps schematically illustrate the neural processing underlying the model.

Figure 3 schematically demonstrates the maximum likelihood calculation for an individual trial. For each confidence probability judgment (c_j) provided by the subject, we can calculate the corresponding decision variable via the inverse Gaussian CDF:

$$\hat{x}_j = \hat{\chi}^{-1}(c_j) = \phi^{-1}(c_j; \hat{\mu}, \hat{k}\hat{\sigma}) \qquad (3)$$

where c_j represents a confidence probability judgment, $\hat{\chi}^{-1}(c_j)$ represents the inverse fitted confidence function, and ϕ^{-1} represents the inverse cumulative Gaussian. The precise probabilistic interpretation of a confidence probability judgment depends on the resolution of the subjective scale provided the subject. Our subjects provided a confidence probability judgment using a scale that had a resolution of 1%. Therefore, when a subject provided a confidence probability judgment of 70%, we set the lower (c_j^{lower}) and upper (c_j^{upper}) bin limits to 69.5 and 70.5%, respectively. Using equation 3, lower (x_j^{lower}) and upper (x_j^{upper}) decision variable limits can be calculated for a given confidence probability judgment, c_j .

As illustrated schematically (Figure 3B), we can then calculate the probability (p_j), which in this context is commonly called the "likelihood", that a decision variable for an individual trial falls in this range using the relationship:

$$p_j = \hat{\Psi}(x_j^{upper}) - \hat{\Psi}(x_j^{lower}) = \phi(x_j^{upper}; s_j + \hat{\mu}, \hat{\sigma}) - \phi(x_j^{lower}; s_j + \hat{\mu}, \hat{\sigma}) \quad (4)$$

where s_j is the stimulus provided on that trial.

Repeating this process for each of the N trials, we can then calculate the log likelihood by simply summing the log of each of the individual trial likelihoods, which can be written as:

$$L\left(\hat{\mu}, \hat{\sigma}, \hat{k}; \vec{c}, \vec{s}\right) = \sum_{j=1}^{N} \log(p_j) \quad (5)$$

We find the maximum likelihood fit by numerically finding the three fit parameters ($\hat{\mu}, \hat{\sigma}, \hat{k}$) that maximize the value of this log likelihood function. This method assumes that the confidence judgment utilizes the same decision variable as the binary decision-making process. Our methods also assume that all processes and mechanisms (e.g., decision boundary, confidence estimation, etc.) are stationary (i.e., constant) across time. This stationarity assumption is standard for psychometric function fits as well. Figure 4 shows example fitted functions.

Data Analysis

To provide a direct comparison of this new fitting method to standard binary forced-choice fitting methods, we fit psychometric curves to the binary data using a maximum likelihood approach (Chaudhuri and Merfeld 2013; Lim and Merfeld 2012). For our forced-choice direction-recognition task, the subject's directional responses are binary (e.g., left or right) and the psychometric function ranges from 0 to 1. A Gaussian distribution was fitted to the data in MATLAB using a generalized linear model using a probit link function. (An example of a psychometric function fit to binary data is shown in Figure 4B.)

The general technique used to fit a psychometric function and a confidence function to the confidence data was described earlier (Fig. 3). To find the maximum likelihood parameter estimates we minimized the negative of the likelihood via a numeric optimization algorithm (MATLAB's fmincon function). The initial value for the confidence-scaling factor ($\hat{k}$) was assumed equal to 1.0; initial values for the psychometric function parameters ($\hat{u},\hat{\sigma}$) were set equal to the values obtained by the GLM fit of the binary forced-choice data.

Human Studies

Basic methods mimicked those used in our earlier studies (Chaudhuri et al. 2013; Valko et al. 2012). Each subject was seated in a racing-style chair with a five-point harness; his/her head was fixed relative to the chair and platform via an adjustable helmet. Each subject wore a pair of noise cancelling earpieces that also provided the ability to communicate with the experimenter. All motions were performed in complete darkness. Subjects performed a binary forced-choice direction-recognition task in response to upright whole-body yaw rotation. Aural white noise began playing in the subject’s earpiece 300 ms before motion commenced and ended when the motion ended. This aural cue was provided to mask any potential directional auditory cues and also informed the subject when a trial began and ended. When the motion and white noise ended, the iPad illuminated and subjects were required to report the motion direction perceived and a confidence probability judgment. Single cycles of sinusoidal acceleration at 1 Hz were used as the motion stimuli. Motion stimuli were generated using a MOOG 6DOF motion platform. There was a pause of at least 3 s between motions. An adaptive sampling procedure – a standard 3-Down/1-Up (3D/1U) staircase using PEST rules (Taylor and Creelman 1967) – was utilized. The initial stimulus amplitude was 4˚/s. Figure 4A shows an example stimulus track for the first 20 trials. There were 100 trials in each experiment.

Subject responses, both the direction responses (i.e., left or right) and quantitative confidence probability judgments having a resolution of 1%, were recorded using an iPad. Before each trial, the iPad backlighting was turned off. When the trial ended, the iPad was automatically illuminated to display sliders (one on the left and one on the right) that ranged from 50% to 100%. The subject tapped on the left side of the iPad to report perceived motion to the left and tapped on the right side to report perceived motion to the right. Subjects could then move the selected slider up/down to indicate their confidence. To avoid biasing the subject’s confidence responses, no indication of slider position was displayed until the subject touched the screen to indicate their confidence (i.e., no initial slider position). The subject's responses – both direction (left/right) and confidence (50% to 100%) – were displayed on the screen. The subject could adjust their response until satisfied; they then tapped a button labeled "Confirm". At the beginning of the testing, the subjects practiced for a few trials to be sure that they understood the task.

These human studies utilized a half-range task in which subjects used a scale between 50% and 100%, inclusive. Confidence probability judgment tasks can be full-range tasks or half-range tasks (Galvin et al. 2003; Lichtenstein et al. 1982). Full-range scales range between 0% and 100%, while half-range scales range between 50% and 100%. For our half-range task, a subject might report that they perceived negative motion and report 84% confidence. For a full range task with the subject asked to report their confidence that the motion was positive, the equivalent response would be a 16% confidence that the motion was positive. To plot, model, and fit the data, we used this mathematical equivalence to convert each half-range confidence rating to a full-range rating.

Subject instructions emphasized that the motion direction would be selected randomly and that the directions of previous motions would have no impact on the next motion direction. Instructions also emphasized that we had no expectation regarding how their confidence assessments would be distributed and that it was important that they report the confidence that they experienced for each specific trial. Subjects were specifically informed that *"…if you are guessing much of the time, this is OK, and if you are very certain much of the time this is OK, too."* Subjects were not ever provided any information regarding their confidence indications. During the initial training that never exceeded 10 practice trials, subjects were informed whether their left/right responses were correct or incorrect. During test sessions, subjects were never informed whether their responses were correct or incorrect.

Four healthy human subjects (2 male, 2 female, 26-34 years old) were each tested on six different days. Informed consent was obtained from all subjects prior to participation in the study. The study was approved by the local ethics committee and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

An author (YY) participated as one of the subjects. Since the computer randomized the motion direction for each trial just before the trial and since the adaptive staircase targets stimuli where all subjects should get about 20% of the trials incorrect, this subject did not have information to guide his binary reports or confidence judgments on each individual trial. More importantly, as noted in the results, this subject's responses did not differ from the other subjects in any noticeable manner.

Simulations

All simulations were performed in MATLAB R2015a (The Mathworks, Inc.) on the Harvard Orchestra computation cluster using parallel IBM BladeCenter HS21 XMs with 3.16 GHz Xeon processors and 8 GB of RAM. These simulations used the same standard adaptive sampling procedure used for the human studies. Specifically, we used a 3-Down/1-Up (3D/1U) staircase having 100 trials. The simulated 3D/1U staircases began at a stimulus level of four. The size of the change in stimulus magnitude was determined using PEST (parameter estimation by sequential testing) rules (Taylor and Creelman 1967).

For all four simulated data sets included herein, the psychometric function, $\Psi(x)=\phi(x;\mu=0.5,\sigma=1)$, and the fitted psychometric function, $\hat{\Psi}(x)=\phi(x;\hat{\mu},\hat{\sigma})$, were modeled as cumulative Gaussians.

For the first simulated data set (1st column of Figs. 8&9), the confidence function was modeled as "well-calibrated" – meaning that the confidence function equaled the psychometric function, $\chi(x)=\Psi(x)=\phi(x;\mu=0.5,\sigma=1)$. For the second simulated data set (2nd column of Figs. 8&9), the confidence function was modeled as "under-confident" with a confidence-scaling factor (k) of 2, yielding a confidence function of $\chi(x)=\phi(x;\mu=0.5,\sigma=2)$. For the last two simulated data sets (3rd and 4th columns of Figs. 8&9), the confidence function was linear – crossing the 0.5 confidence level with a bias of 0.5 $(\mu=0.5)$ and with a slope of 0.1443 ($m=0.1443$), $\chi(x)=m(x-\mu)+0.5=0.1443x+0.428$, which yielded confidence saturations at -2.96 ("zero") and +3.96 ("one").

For the first three simulated data sets (1st, 2nd, and 3rd columns of Figs. 8&9), we fit the data by modeling the fitted confidence function as the 3-parameter Gaussian CDF introduced as Equation 3, $\hat{\chi}(x)=\phi(x,\hat{\mu},\hat{k}\hat{\sigma})$. For the last simulated data set (4th column of Figs. 8&9), we fit the data by modeling the fitted confidence function as a linear model of confidence, $\hat{\chi}(x)=\hat{m}(x-\hat{\mu})+0.5$, which matched the form of the simulated confidence function.

Two simulation data sets included additive noise. For the simulations shown in 3rd column of Figs. 8&9, the noise distribution was modeled as zero-mean uniform noise having width of 20% (-10% to +10%), $U(-0.1,+0.1)$. We intentionally chose a large noise range to demonstrate the small impact of such confidence noise. The relative stability of human confidence ratings suggest that this simulated noise level overestimates the actual contributions of confidence noise. For context, recognize that this noise distribution means that if the confidence function yielded a confidence rating of 70%, that the noise would lead to a report of between 60% and 80%, essentially yielding roughly 3 functional confidence bins between 50% (just guessing) and 100% (certain). We chose uniform noise because we could strictly control the noise range without any impact to the noise distribution. This was important because confidence must stay in the range 0% to 100%. To keep the noise zero-mean when the confidence function yielded a confidence of greater than 90% (or less than 10%), the limits on the noise distribution were reduced to keep the confidence judgment in the range of 0 to 100%. For example, if the simulation yielded a mean confidence of 94% for an individual trial, the noise distribution was modeled as zero-mean uniform with a range of -6% to 6% (i.e., $U(-0.06,+0.06)$) for that trial. This had no impact on most trials, as confidence usually was lower than 90%. For the simulations shown in the 4th column of Figs. 8&9, the noise distribution was again modeled as zero-mean uniform noise but with a width of 10% (-5% to +5%), $U(-0.05,+0.05)$.

Results

Human Studies

Fitted psychometric function $(\hat{\mu},\hat{\sigma})$ and confidence scaling $(\hat{k})$ parameters for each of our four subjects for yaw rotation about an earth-vertical rotation axis are shown in Figs. 5 (mean) and 6 (standard deviation); parameter fits are plotted versus the number of trials in increments of 5 trials starting at the 15th trial. (To demonstrate raw performance for individual test sessions, Appendix Fig. B1 presents the parameter fits for each of the six individual tests for each subject.) As described in the methods, all parameter estimates are determined using maximum likelihood methods.

Consistent with previous studies utilizing adaptive procedures (e.g., Chaudhuri and Merfeld 2013; Garcia-Perez and Alcala-Quintana 2005), the conventional estimates of the width of the psychometric function ($\hat{\sigma}$) took between 50 and 100 trials to stabilize (Figs. 5A-D, black curves). More specifically, using these conventional psychometric methods, the estimated width parameter ($\hat{\sigma}$) was significantly lower after 20 trials than after 100 trials (repeated measures ANOVA, N=4 subjects, p=0.011).

In contrast, estimates of the width parameter ($\hat{\sigma}$) using our confidence fit technique required fewer than 20 trials to reach stable levels (Figs. 5A-D, red curves). Specifically, the width parameter ($\hat{\sigma}$) estimated using confidence probability judgments was not significantly different after 20 trials than for 100 trials (repeated measures ANOVA, N=4 subjects, p=0.251). Furthermore, the estimated width parameter after 20 trials using confidence probability judgments was not significantly different from the estimated width parameter after 100 trials using conventional psychometric fit methods (repeated measures ANOVA, N=4 subjects, p=0.907).

Furthermore, the parameter estimates obtained using conventional psychometric fits (Fig. 6, black traces) were more variable than the fits obtained using our CSD model (Fig. 6 gray traces). In fact, the precision of the psychometric width estimate using the confidence model was about the same after 20 trials (average standard deviation of

0.124 across subjects) as the conventional psychometric fit estimate after 100 trials (0.129).

The estimates of the shift of the psychometric functions ($\hat{\mu}$) showed a qualitatively similar pattern; the estimates that utilized confidence reached stable levels a little sooner and were more precise than the estimates provided by the conventional analysis. We also note that three of our subjects seemed well calibrated (Figs. 5E, F, and G) with fitted confidence-scaling factors near 1, while the second subject had a fitted confidence-scaling factor near 2 (Fig. 5H), suggesting substantial under-confidence.

Simulations

We also simulated tens of thousands of test sessions to test the confidence fit procedures more thoroughly. The simulations were designed to mimic the human studies with the obvious difference being that we defined the simulated psychometric ($\Psi(x)$) and confidence ($\chi(x)$) functions. Since we knew these simulated functions, this allowed us to quantify parameter fit accuracy. For all simulated data sets, we fit the conventional binary forced-choice data and compared and contrasted these fits with the CSD fits Histograms show fitted parameters after 20 (Fig 7, left three columns) and 100 (Fig. 7, right three columns) trials for 10,000 simulations. After as few as 20 trials, the CSD fit parameters demonstrated relatively tight distributions (Fig. 7B, C) in comparison to the binary fits that show ragged distributions (Fig 7A). After 100 trials, the binary fit parameters demonstrated relatively tight distributions (Fig. 7D) that mimicked those found for the CSD fit parameters after 20 trials (Fig. 7B, C). The CSD fit parameters after 100 trials (Fig. 7E, F) demonstrated higher precision (i.e., lower variance) than the binary fit parameters after 100 trials (Fig. 7D). (APPENDIX Fig. B2 presents similar histograms for 100 trials for the other two simulation data sets.)

Mimicking the format previously used for the human data (Figs. 5 and 6); simulation parameter fits are plotted versus the number of trials in increments of 5 trials starting at the 15th trial. The black curves in Figs. 8 and 9 show the fitted psychometric function parameters for the binary forced choice data; the red (Fig. 8) and gray (Fig. 9) curves show the fitted psychometric and confidence function parameters fit using the CSD model. (To provide direct quantitative comparisons, Appendix B summarizes data from all simulations in tabular form.)

The simulated data (1st column of Figs. 8 & 9 and 2nd row of APPENDIX B Tables B1-3) show that the CSD model yielded fit parameters that accurately matched those simulated when the simulated subject's confidence was well calibrated ($k=1$), where "well calibrated" means that the subject's confidence matches the psychometric function, $\chi(x)=\phi(x;\mu=0.5,k\sigma=1)$. Even when the subject's confidence was not well calibrated ($k=2$), the confidence fit parameters matched the three confidence function parameters well (2nd column of Figs. 8 & 9 and 3rd row of APPENDIX B Tables B1-3). In fact, except that the fitted confidence-scaling factor ($\hat{k}$) settles near a value of 2 (Fig. 8F) instead of 1 (Fig. 8E), the average psychometric parameter estimates for an under-confident subject appeared nearly the same as for a well-calibrated subject. Indeed, the fitted psychometric width parameter ($\hat{\sigma}$) demonstrated a lower standard deviation for an under-confident subject than for a well-calibrated subject (see APPENDIX B).

To demonstrate robustness, we utilized the same Gaussian confidence fit model (Eq. 2) while simulating a confidence model that differed from the Gaussian confidence fit model in two ways. First, we modeled the confidence function as a linear function (slope of 0.1445; i.e., $\sigma = 2$) instead of a cumulative Gaussian. And, secondly, we added zero-mean uniform noise, $U(-0.1, 0.1)$, to the simulated confidence response. Despite these differences, the confidence fit of these simulated data mimics the earlier confidence fits well (3rd column of Figs. 8&9 and 4th row of APPENDIX B Table B1-3). The primary difference is that the parameter fit precision was not as good as for the first two simulation sets described above but was still better than for the conventional fits. For example, despite the severe noise (-10% to +10%), the fit precision for the width parameter ($\hat{\sigma}$) after 20 trials utilizing confidence matched the fit precision after about 50 trials using conventional analyses.

Finally, to demonstrate the flexibility of the confidence fit technique, we model the same linear confidence function from the previous paragraph, but we now add less extreme zero-mean uniform noise levels ($U(-0.05, 0.05)$) and fit a linear confidence function that mimics the linearity of the true confidence function used for these simulations. The fit accuracy and precision were very good (4th column of Figs. 8 & 9 and 5th row of APPENDIX B Tables B1-3) – demonstrating that the fitted psychometric function and confidence function need not be similar in form. Appendix Table B4 presents some conventional confidence metrics, including goodness of fit parameters.

Discussion

This paper introduces a new confidence signal detection (CSD) model (Fig. 1) and then uses this model to develop a confidence analysis technique (Fig. 3) that utilizes confidence probability judgments to help yield more efficient psychometric function fits. The primary novelty of the new technique is the introduction of a confidence function - alongside confidence probability judgments - to help improve psychometric function fit efficiency. The introduction of a fitted confidence function yields two benefits. First, it provides a way to incorporate confidence probability judgments into a psychometric fit procedure. Second, while it does not require that the confidence function differ from the psychometric function, it allows these two functions to differ from one another.

We assumed that the binary decision variable is used to determine confidence probability judgments. Of course, if confidence does not correlate in some manner with the sampled decision variable used to make a decision, this would lead to erroneous results but empirical data presented suggest that, at least for yaw rotation thresholds, this does not appear to be a substantive concern.

We specifically note that while we utilized confidence probability judgments and a confidence function, similar benefits may accrue by replacing (a) the confidence function and confidence recordings with a magnitude estimation function accompanied by magnitude estimation recordings (e.g., I rotated +3˚) or other analogous perceptual functions and associated recordings or (b) the confidence probability judgments by another confidence assay accompanied by an appropriate model of the confidence assay. Such potential benefits would need to be investigated theoretically and empirically via behavioral studies and simulations as provided herein for confidence probability judgments.

In this paper, we focus nearly exclusively on improving the efficiency of psychometric parameter fits, but the fitting technique shown herein also calculates a confidence-scaling factor, which may independently prove useful for studies of confidence calibration. More generally, the confidence modeling approach described herein (Fig. 1) may benefit studies of confidence calibration, but such topics require separate study.

Human Studies

Our human experimental data showed that stable psychometric function parameters could be estimated in as few as 20 trials (Figs. 5 and 6). In fact, these human data demonstrated no significant differences between the psychometric function parameters estimated after 20 trials using confidence data and the same parameters estimated after 100 trials using conventional methods. These findings suggest that the number of trials required to estimate psychometric fit parameters could be reduced by roughly 80% simply by collecting a confidence probability judgments and utilizing the confidence information in the manner shown.

Not one of these four subjects had performed an experiment utilizing a confidence probability judgment prior to our studies of confidence and none were provided any feedback, except during the initial practice that consisted of 10 trials when subjects were informed whether their responses were correct or incorrect. No specific feedback regarding confidence was ever provided during the course of the study. Despite intentionally limiting the training and feedback, each subject yielded a coherent data set across the six test sessions. It is worth noting that all four subjects were experienced observers; one of the subjects was an author (YY), but there was nothing noteworthy about his data as he was not the subject whose confidence-scaling factor was near 2.

For our task, we minimize the potential impact of after-effects from the previous trial (Coniglio and Crane 2014; Crane 2012a; b) by requiring at least 3 seconds between the end of a stimulus and the start of the next stimulus (Chaudhuri et al. 2013). Therefore, providing confidence did not require much, if any, additional time to complete each trial than a conventional binary forced-choice task for our direction-recognition task, but this may not be true for all other tasks. Nonetheless, an 80% improvement in efficiency may be well worth a little more time on each trial even for those tasks that do not include such a mandatory inter-trial interval.

We also want to note at least a possible concern/limitation. While we did not attempt to test children for this study, it seems highly likely that forced-choice tasks can be taught more readily to children and probably even some adults (e.g., elderly patients) than the confidence probability judgment task we utilized. Some simple form of training with feedback may prove beneficial but we intentionally avoided any feedback for this initial study. We also specifically note that a better understanding of confidence would be required for maximal utilization of this new CSD psychometric analysis.

For example, the benefit of confidence training for perceptual tasks needs to be studied. For knowledge tasks, investigations show that simple and short training sessions that provide direct confidence calibration information can lead to improved calibration (Lichtenstein and Fischhoff 1980; Stone and Opel 2000) and that such improvement might generalize to tasks beyond the one calibrated (Lichtenstein and Fischhoff 1980). If generalized calibration occurs, such training could help calibrate confidence prior to testing as part of the process by which subjects are taught the requisite tasks; this may be particularly useful for clinical testing. Alongside existing measures (e.g., the Brier score and its calibration and resolution components (Yates 1990)), our confidence-scaling factor might be useful to help determine if confidence calibration is an individual trait and to help train confidence calibration. Beyond improving psychometric parameter estimation, a specific benefit of our confidence modeling approach may be that it only requires about 20 to 50 trials (e.g., Figs. 5 and 6) to provide relatively stable calibration feedback. This is substantially fewer trials than the hundreds of trials used before (Lichtenstein and Fischhoff 1980; Stone and Opel 2000) because the calibration plots and the Brier score partitions used previously are sensitive to sample size (Lichtenstein and Fischhoff 1980).

Simulations

Simulations confirmed that this new fitting technique yields accurate psychometric parameter estimates and also confirmed a marked efficiency improvement. Even simulations that assumed (a) a confidence model that was not matched by the fitting model and (b) large additive confidence noise - likely greater than actual confidence noise - yielded psychometric function parameter estimates that matched the simulated psychometric function much more efficiently than conventional psychometric fits (i.e., 20 trials compared to 50-100 trials). If the assumed confidence function is wrong, it could lead to biased predictions. However, simulations show that even assuming the wrong confidence function did not cause major difficulties.

Simulations showed that this new fitting method works well for psychometric functions that range from 0 to 1. As noted above, we also verified experimentally that this fitting method works well for a specific vestibular direction-recognition task, but we did not verify experimentally that this new fitting technique works for other tasks/modalities. Nonetheless simulation results should be generally applicable to all tasks yielding psychometric functions that range from 0 to 1. Furthermore, there is no fundamental reason to believe that this technique cannot be applied to other tasks – like detection tasks (i.e., yes/no tasks) or to two-alternative forced choice detection tasks where the subject identifies the interval (or location) when (or where) the signal occurred – having different psychometric ranges (e.g., 0.5 to 1). But we have not yet determined the extent to which our findings generalize to psychometric functions having other ranges.

In closing, we have developed a new CSD model (Fig. 1) and then used this new model to develop a confidence analysis technique (Fig. 3) that utilizes confidence probability judgments that can reduce the number of trials needed to provide good psychometric parameter estimates. Human studies (Figs. 5 and 6) using a direction-recognition task and a psychometric function that varies from 0 to 1 and extensive simulations (Figs. 7-9) suggest that about 20 trials using this new confidence fit method can match the performance typically achieved only after about 100 trials using conventional fit methods.

Grants

This research was supported by NIH/NIDCD R01-DC04158 and R56-DC12038.

Disclosures

The authors declare no competing financial interests.

Author Contributions

The first author performed the experiments and analyzed all data. Both authors contributed to experimental design, contributed to theoretic and simulated results, and wrote the manuscript.

Acknowledgements

We appreciate the participation of our anonymous subjects. We thank Sho Chaudhuri, Torin Clark, Raquel Galvan-Garza, Faisal Karmali, Koeun Lim, and Wei Wang for helpful comments on earlier manuscript drafts. We also thank Bob Grimes and Wangsong Gong for technical support.

Appendix A: Flow chart for confidence fit process

Specific Gaussian confidence fit process	General confidence fit process
A. Experimentally record a confidence probability judgment, c_j , for each of n stimuli, s_j , that explicitly incorporates a binary (i.e., two-alternative) decision.	**A**. Experimentally record a confidence rating, c_j , for each of n stimuli, s_j , that explicitly or implicitly includes an m-alternative decision.
B. Choose a cumulative Gaussian: $\hat{\Psi}(x) = \phi(x;\hat{\mu},\hat{\sigma})$ as the psychometric function, $\hat{\Psi}(x)$, to fit the data.	**B**. Via empiric or theoretic means, choose an appropriate psychometric function, $\hat{\Psi}(x)$, to fit the data.
C. Choose a cumulative Gaussian whose standard deviation differs from the psychometric function via a fitted scalar value, $\hat{k}$: $\hat{\chi}(x) = \phi(x;\hat{\mu},\hat{k}\hat{\sigma})$ as the confidence function, $\hat{\chi}(x)$, to fit the data.	**C**. Via empiric or theoretic means, choose an appropriate confidence function, $\hat{\chi}(x)$, to fit the data. (Generally, this confidence function can even differ in form from the psychometric function.)
D. For each confidence probability judgment, c_j , set the upper(c_j^{upper}) and lower bin limits(c_j^{lower}).	**D**. For each confidence rating, c_j , set (or determine as part of the fit procedure) the upper and lower bin limits.
E. Choose initial values (presumably near the expected fit values) for each of the three fit parameters $\hat{\mu}, \hat{k},$ and. $\hat{\sigma}$.	**E**. Choose initial values (presumably near the expected fit values) for each of the parameters to be fit, ($\hat{\vec{\theta}}^{initial}$).
F. For each confidence probability judgment calculate the upper and lower limit on the decision variable using the inverse of the fitted confidence function, $\hat{\chi}^{-1}(c)$: $x_j^{upper} = \phi^{-1}(c_j^{upper}, 0, \hat{k}\hat{\sigma})$ $x_j^{lower} = \phi^{-1}(c_j^{lower}, 0, \hat{k}\hat{\sigma})$	**F**. For each confidence rating calculate the upper and lower limit on the decision variable using the inverse of the fitted confidence function, $\hat{\chi}^{-1}(c)$: $x_j^{upper} = \hat{\chi}^{-1}(c_j^{upper})$ $x_j^{lower} = \hat{\chi}^{-1}(c_j^{lower})$
G. With this range for the decision variables for the given stimulus (s_j), we can calculate the probability of this specific confidence probability judgment given the fitted psychometric function: $p_j = \phi(x_j^{upper}, s_j + \hat{\mu}, \hat{\sigma}) - \phi(x_j^{lower}, s_j + \hat{\mu}, \hat{\sigma})$	**G**. With this range for the decision variables for the given stimulus (s_j), we can calculate the probability of this specific confidence probability judgment given the fitted psychometric function: $p_j = \hat{\Psi}(x_j^{upper}) - \hat{\Psi}(x_j^{lower})$
H. Repeat steps F and G n times (once for data from each of n trials) and calculate a log likelihood function by summing the logarithm of each of the n probability values: $L\left(\hat{\mu},\hat{\sigma};\vec{c},\vec{s}\right) = \sum_{j=1}^{n} \log(p_j)$.	**H**. Repeat steps F and G n times (once for data from each of n trials) and calculate an appropriate cost function. $C\left(\hat{\vec{\theta}};\vec{c},\vec{s}\right) = g(p_j)$
I. Repeat steps F through H while varying $\hat{\mu}, \hat{k},$ and. $\hat{\sigma}$ to maximize the log likelihood function.	**I**. Repeat steps F through H while varying the fit parameters ($\hat{\vec{\theta}}$) to optimize the cost function.

Appendix B: Simulated fit parameters

Tables included in this appendix summarize results from 10,000 simulated test sessions for direct quantitative comparisons. These fit parameters are the same as shown graphically in Figs. 7&8. The last row of Tables B1 and 3 quantitatively presents the fitted psychometric parameters using the conventional binary methods; these were graphically presented in Figs. 7&8 as black curves. The 1st to the 4th rows of Tables B1-3 quantitatively present the fitted psychometric and confidence scaling parameters found using the CSD fit; these were graphically presented in Figs. 7&8 as red or gray curves. Specifically, the 1st row of Tables B1-3 quantitatively presents fit parameters for a well-calibrated subject ($k=1$) when both confidence and confidence fit functions were cumulative Gaussians (1st column of Figs 7&8). The 2nd row of Tables B1-3 quantitatively presents fit parameters for an under-confident subject ($k=2$) when both confidence and confidence fit functions were cumulative Gaussians (2nd column of Figs 7&8). The 3rd row of Tables B1-3 quantitatively presents fit parameters an under-confident subject when the confidence function is linear, $\chi(x)=m(x-\mu)+0.5=0.1443x+0.428$, with added zero-mean uniform noise ($U(-0.1,+0.1)$), and the confidence fit function was a cumulative Gaussian (3rd column of Figs 7&8). The 4th row of Tables B1-3 present fit parameters for an under-confident subject with the same linear confidence function with added zero-mean uniform noise ($U(-0.05,+0.05)$) when the confidence fit function was linear, $\hat{\chi}(x)=\hat{m}(x-\hat{\mu})+0.5$ (4th column of Figs 7&8). Table B4 shows some conventional parameters that summarize different characteristics of the fit, including goodness of fit.

Table B1. Fitted width parameter ($\hat{\sigma}$).

Trials	20	40	60	80	100
Fig. 7, 1st col.	0.926 (0.213)	0.960 (0.163)	0.973 (0.139)	0.977 (0.121)	0.983 (0.111)
Fig. 7, 2nd col.	0.946 (0.188)	0.971 (0.144)	0.980 (0.125)	0.984 (0.111)	0.987 (0.102)
Fig. 7, 3rd col.	1.053 (0.279)	1.041 (0.218)	1.045 (0.189)	1.050 (0.166)	1.058 (0.153)
Fig. 7, 4th col.	0.955 (0.203)	0.984 (0.156)	0.995 (0.134)	0.999 (0.119)	1.003 (0.109)
Binary	0.588 (0.484)	0.841 (0.340)	0.914 (0.257)	0.944 (0.209)	0.959 (0.179)

Table B2. Fitted confidence scaling factor ($\hat{k}$)

Trials	20	40	60	80	100
Fig. 7, 1st col.	1.102 (0.234)	1.045 (0.137)	1.029 (0.107)	1.021 (0.090)	1.016 (0.079)
Fig. 7, 2nd col.	2.189 (0.389)	2.089 (0.245)	2.058 (0.194)	2.043 (0.164)	2.034 (0.146)
Fig. 7, 3rd col.	1.752 (0.464)	1.814 (0.360)	1.845 (0.308)	1.859 (0.272)	1.868 (0.248)
Fig. 7, 4th col.	0.148 (0.022)	0.148 (0.018)	0.147 (0.015)	0.147 (0.014)	0.146 (0.013)

Table B3. Fitted bias parameter ($\hat{\mu}$)

Trials	20	40	60	80	100
Fig. 7, 1st col.	0.499 (0.255)	0.494 (0.175)	0.496 (0.143)	0.497 (0.123)	0.497 (0.111)
Fig. 7, 2nd col.	0.505 (0.237)	0.500 (0.167)	0.500 (0.137)	0.500 (0.120)	0.499 (0.107)
Fig. 7, 3rd col.	0.486 (0.283)	0.468 (0.193)	0.469 (0.158)	0.471 (0.136)	0.473 (0.122)
Fig. 7, 4th col.	0.503 (0.249)	0.496 (0.173)	0.498 (0.143)	0.498 (0.124)	0.497 (0.110)
Binary	0.588 (0.529)	0.514 (0.308)	0.501 (0.232)	0.498 (0.191)	0.498 (0.166)

Table B4. Conventional confidence metrics for human and simulated data

	BS	REL	RES	UNC	Deviance Binary	Deviance CSD
Subject 1	0.150 (0.015)	0.081 (0.013)	0.181 (0.016)	0.249 (0.001)	92.6 (8.7)	789.0 (40.8)
Subject 2	0.127 (0.021)	0.074 (0.017)	0.192 (0.016)	0.246 (0.003)	75.3 (10.1)	752.6 (44.8)
Subject 3	0.145 (0.020)	0.081 (0.011)	0.185 (0.019)	0.249 (0.001)	93.6 (9.5)	791.3 (31.0)
Subject 4	0.147 (0.025)	0.076 (0.014)	0.176 (0.012)	0.247 (0.004)	79.2 (12.8)	821.0 (20.6)
Fig. 7, 1st col.	0.135 (0.019)	0.095 (0.015)	0.207 (0.015)	0.247 (0.004)	79.3 (8.0)	764.6 (37.4)
Fig. 7, 2nd col.	0.141 (0.015)	0.091 (0.012)	0.198 (0.016)	0.247 (0.004)	79.3 (8.0)	807.9 (18.2)
Fig. 7, 3rd col.	0.152 (0.015)	0.093 (0.013)	0.188 (0.017)	0.247 (0.004)	79.3 (8.0)	830.7 (22.7)
Fig. 7, 4th col.	0.150 (0.014)	0.092 (0.013)	0.190 (0.017)	0.247 (0.004)	79.3 (8.0)	788.0 (21.7)

Brier Score (BS) and its three decomposed components; Reliability (REL), Resolution (RES), and Uncertainty (UNC) are shown. For definitions and descriptions, see (Brier 1950; Murphy 1973; 1972). We use the formulation commonly used today, which results by dividing Brier's original formulation by two. Deviance for each of two fits are also shown. Mean (and standard deviation) across 6 trials for each subject and across 10,000 simulations for each simulated data set are provided.

Figure Legends

Fig. 1. Relationship between decision variables, psychometric functions ($\Psi(x)$), and confidence functions ($\chi(x)$) in our confidence signal detection (CSD) model. *A:* The stimulus for this example is well controlled having an amplitude of +1.0 with little variation, so the objective probability density function (PDF) is a delta function. *B:* A signal detection model assumes additive noise. For this example, Gaussian noise having zero-mean and a standard deviation of 1 is added to the stimulus of +1.0 and leads to the subjective PDF shown. The dotted vertical line at zero represents a decision boundary. If a sampled decision variable falls to the right of the decision boundary, represented by the gray area, the subject decides positive. If the sampled decision variable falls to the left, the subject decides negative. For this example, 84% of the decision variables lead to the subject deciding positive. *C:* The asterisk, located at (1, 0.84) represents the example data point illustrated in the previous panel. When this process is repeated for a variety of different stimulus levels, it yields a psychometric function, $\Psi(x)$ (black curve). *D:* Similarly, a relationship between confidence and the stimulus for an individual trial can be represented by a confidence function ($\chi(x)$). The psychometric function represents average subject performance, and confidence is defined as well-calibrated when confidence matches average subject performance (Bjorkman et al. 1993; Ferrell 1995; Keren 1991; Lichtenstein et al. 1982; Stankov et al. 2012). Therefore, well-calibrated confidence matches the psychometric function ($\chi(x) = \Psi(x)$) and is plotted as the solid curve. Also shown are confidence functions that represent over-confidence (dashed) and under-confidence (dotted).

Fig. 2. Schematic representation of pertinent neural processing. For the experimental investigations herein, the stimulus is angular velocity (ω), which is transduced and processed via pertinent perceptual mechanism to yield a perceptual representation of angular velocity ($\hat{\omega}$); these representations could readily be generalized to any perceptual process. This perceptual signal may be further filtered (e.g. Merfeld et al. 2015) to yield a decision-variable (*d*) used both to make the binary decision – by comparing the decision variable to the decision boundary – and via additional neural manipulations to yield confidence (*c*) in the decision. The latter confidence calculations may be performed "correctly" [i.e., $c = \chi(d) = \phi(d, \mu, \sigma)$], yielding well-calibrated data. Alternatively, these confidence calculations could be miscalibrated [i.e., $c = \chi(d) = \phi(d, \mu, k\sigma)$], where $k \neq 1$, leading to overconfidence ($k < 1$) or underconfidence ($k > 1$). Furthermore, as we will evaluate via simulations, the confidence function may be independent of the noise distribution that defines the psychometric function (i.e., confidence function is not Gaussian for this application).

Fig. 3. Illustration of how confidence probability judgments from individual trials contribute to a maximum likelihood psychometric function fit. *A:* Given a confidence probability judgment, we can use the inverse fitted confidence function ($\hat{\chi}^{-1}(c_j)$) to calculate a modeled decision variable to accompany that judgment. More specifically, given upper and lower limits to a confidence probability judgment (dashed horizontal lines), we can use the inverse fitted confidence function to calculate the corresponding upper and lower decision variable limits (dashed vertical lines). *B:* Given the estimated decision variable range shown by the dashed vertical lines, we can calculate the

probability that the given stimulus (s_j) and psychometric function noise model would yield that confidence probability judgment. Two examples are illustrated. The light curve shows the decision variable PDF for the stimulus having an amplitude of +1.0 shown in Fig. 1B and the light shaded area represents the probability of the confidence probability judgment for a +1.0 stimulus. The dark curve shows the PDF for a stimulus having an amplitude of -1.0, and the dark shaded area represents the probability for a -1.0 stimulus. High confidence that the motion is positive is much more probable (i.e., much more likely) for the +1 stimulus than for the -1 stimulus.

Fig. 4. Example fits for a human test. *A:* Example stimulus track, including confidence probability judgments, is shown for first 20 trials. Upward-pointing gray triangles and downward-pointing black triangles represent rightward and leftward trials, respectively. *B:* Following 20 binary forced-choice trials, a conventional psychometric function (black curve), $\hat{\Psi}(x)=\phi(x;\hat{\mu}=0.05,\hat{\sigma}=0.95)$, was fit to the binary forced-choice data points shown. *C:* Given the same 20 trials with confidence probability judgments, a psychometric function (black curve), $\hat{\Psi}(x)=\phi(x;\hat{\mu}=0.19,\hat{\sigma}=0.91)$, and a confidence function (gray curve), $\hat{\chi}(x)=\phi(x;\hat{\mu}=0.19,\hat{k}\hat{\sigma}=1.43)$, were simultaneously fit to the confidence data. All example data are from one of the human data sets (Fig. 5, 4th column) presented herein. For comparison, the fitted psychometric function determined after 100 binary forced-choice trials using conventional methods, $\hat{\Psi}(x)=\phi(x;\hat{\mu}=0.33,\hat{\sigma}=0.59)$, is also shown via dashed lines on panels b and c. Half-scale (50% to 100%) probability judgments provided by subjects have been converted to full-scale (0 to 100%) judgments as described in Methods.

Fig. 5. Summary of human psychometric parameter estimates as trial number increases. Each column represents fitted parameters for one subject. Top row (A-D) shows average fitted psychometric width parameter ($\hat{\sigma}$). Middle row (E-H) shows average fitted confidence scaling factor ($\hat{k}$). Bottom row (I-L) shows average fitted psychometric function bias ($\hat{\mu}$). Thick black curves show average psychometric parameter estimates calculated using conventional forced-choice analyses. Thick red curves show average parameter estimates determined by fitting confidence probability judgment data. Errors bars (thin gray curves and thin red curves, respectively) represent standard deviation of parameter estimates.

Fig. 6. Standard deviation of human psychometric parameter estimates as trial number increases. Each column represents fitted parameters for one subject in the same order as Fig. 5. Top row (A-D) represents the standard deviation of the fitted psychometric width parameter ($\hat{\sigma}$). Bottom row (E-H) represents the fitted psychometric function bias ($\hat{\mu}$). Black curves show standard deviation of psychometric parameter estimates calculated using conventional forced-choice analyses. Gray curves show standard deviation of parameter estimates determined via our CSD model fit.

Fig 7. Parameter distributions show parameter estimates for 10,000 simulated experiments with 20 and 100 trials. The columns from left to right represent the fitted psychometric width parameter ($\hat{\sigma}$), the fitted confidence scaling factor ($\hat{k}$) and the fitted psychometric function bias ($\hat{\mu}$) as shown on the x-axis at bottom. Top row (A and D) represents fitted parameters of conventional binary forced-choice parameter estimates. Middle row (B and E) represents fitted parameters estimates determined via our CSD

model fit for a well-calibrated subject ($k=1$). Bottom row (C and F) represents fitted parameters estimates determined via our CSD model fit for an under-confident subject ($k=2$). The solid black line shows the actual parameter value (i.e. $\mu=0.5$ or $\sigma=1$), the solid gray line shows the mean of fitted parameters, and the dashed gray lines indicate standard deviation on each side of the mean.

Fig. 8. Summary of simulation parameter estimates as trial number increases. As illustrated via insets, each column represents different simulated combinations of the confidence function (red solid curves) and the fitted confidence function (red dashed curves). First column shows a well-calibrated subject ($k=1$) when both confidence and confidence fit functions are cumulative Gaussians. Second column shows an under-confident subject ($k=2$) when both confidence and confidence fit functions are cumulative Gaussians. Third column shows an under-confident subject when the confidence function is linear, $\chi(x)=m(x-\mu)+0.5=0.1443x+0.428$, with added zero-mean uniform noise ($U(-0.1,+0.1)$), and the confidence fit function is a cumulative Gaussian. Fourth column shows an under-confident subject with the same linear confidence function with added zero-mean uniform noise ($U(-0.05,+0.05)$) when the confidence fit function is linear, $\hat{\chi}(x)=\hat{m}(x-\hat{\mu})+0.5$. Top row (A-D) shows fitted psychometric width parameter ($\hat{\sigma}$). Middle row (E-G) shows fitted confidence-scaling factor ($\hat{k}$) or (H) fitted slope of confidence function. Bottom row (I-L) shows fitted psychometric function bias ($\hat{\mu}$). Thick black curves show average conventional forced-choice parameter estimates, which are identical for all conditions. Thick red curves show average parameter estimates determined by fitting confidence probability

judgments. Errors bars (thin gray curves and thin red curves, respectively) represent standard deviation of parameter estimates.

Fig. 9. Standard deviation of simulation parameter estimates as trial number increases. Each column represent the same conditions as Fig. 8. Top row (A-D) represents the fitted psychometric width parameter ($\hat{\sigma}$). Bottom row (E-G) represents the fitted psychometric function bias ($\hat{\mu}$). Black curves show standard deviation of conventional forced-choice parameter estimates, which are identical for all conditions. Gray curves show standard deviation of parameter estimates determined via our CSD model fit.

Fig. B1. Human psychometric width parameter ($\hat{\sigma}$), confidence scaling factor ($\hat{k}$) and, bias parameter ($\hat{\mu}$) estimates as trial number increases for each subject for each of 6 test sessions. Each column shows fitted parameters for one subject. Top row (A-D) shows fitted psychometric width parameter using conventional forced-choice analyses. Second row (E-H**)** shows fitted psychometric width parameter for CSD model fit. Third row (I-L) shows fitted confidence scaling factor for CSD model fit. Fourth row (M-P) shows fitted psychometric function bias using conventional forced-choice analyses. Bottom row (Q-T) shows fitted psychometric function bias for CSD model fit.

Fig. B2. Parameter distributions show parameter estimates for 10,000 simulated experiments with 20 and 100 trials in the same order as figure 6. Top row (A and C) represents fitted parameters estimates determined via our CSD model fit for an under-confident subject when the confidence function is linear, $\chi(x) = m(x-\mu) + 0.5 = 0.1443x + 0.428$, with added zero-mean uniform noise ($U(-0.1,+0.1)$), and the confidence fit function is a cumulative Gaussian. Bottom row (B

and D) represents fitted parameters estimates determined via our CSD model fit for an under-confident subject with the same linear confidence function with added zero-mean uniform noise ($U(-0.05,+0.05)$) when the confidence fit function is linear, $\hat{\chi}(x) = \hat{m}(x - \hat{\mu}) + 0.5$. The solid black line shows the actual parameter value (i.e. $\mu = 0.5$ or $\sigma = 1$), the solid gray line shows the mean of fitted parameters, and the dashed gray lines indicate standard deviation on each side of the mean.

References

Agresti A. *An introduction to categorical data analysis*. Wiley New York, 1996.

Balakrishnan JD. Decision processes in discrimination: fundamental misrepresentations of signal detection theory. *Journal of experimental psychology Human perception and performance* 25: 1189-1206, 1999.

Baranski JV, and Petrusic WM. The calibration and resolution of confidence in perceptual judgments. *Perception & psychophysics* 55: 412-428, 1994.

Björkman M. Internal cue theory: Calibration and resolution of confidence in general knowledge. *Organizational behavior and human decision processes* 58: 386-405, 1994.

Bjorkman M, Juslin P, and Winman A. Realism of confidence in sensory discrimination: the underconfidence phenomenon. *Perception & psychophysics* 54: 75-81, 1993.

Brier GW. Verification of forecasts expressed in terms of probability. *Monthly weather review* 78: 1-3, 1950.

Chaudhuri SE, Karmali F, and Merfeld DM. Whole body motion-detection tasks can yield much lower thresholds than direction-recognition tasks: implications for the role of vibration. *Journal of neurophysiology* 110: 2764-2772, 2013.

Chaudhuri SE, and Merfeld DM. Signal detection theory and vestibular perception: III. Estimating unbiased fit parameters for psychometric functions. *Exp Brain Res* 225: 133-146, 2013.

Coniglio AJ, and Crane BT. Human Yaw Rotation Aftereffects with Brief Duration Rotations Are Inconsistent with Velocity Storage. *Journal of the Association for Research in Otolaryngology : JARO* 2014.

Crane BT. Fore-aft translation aftereffects. *Exp Brain Res* 219: 477-487, 2012a.

Crane BT. Roll aftereffects: influence of tilt and inter-stimulus interval. *Exp Brain Res* 223: 89-98, 2012b.

Drugowitsch J, Moreno-Bote R, and Pouget A. Relation between belief and performance in perceptual decision making. *PLoS ONE* 9: e96511, 2014.

Ferrell WR. A model for realism of confidence judgments: Implications for underconfidence in sensory discrimination. *Perception & psychophysics* 57: 246-254, 1995.

Ferrell WR, and McGoey PJ. A model of calibration for subjective probabilities. *Organizational Behavior and Human Performance* 26: 32-53, 1980.

Fleming SM, and Dolan RJ. The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367: 1338-1349, 2012.

Galvin SJ, Podd JV, Drga V, and Whitmore J. Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychon Bull Rev* 10: 843-876, 2003.

García-Pérez MA, and Alcalá-Quintana R. Interval bias in 2AFC detection tasks: sorting out the artifacts. *Atten Percept Psychophys* 73: 2332-2352, 2011.

García-Pérez MA, and Alcalá-Quintana R. Shifts of the psychometric function: Distinguishing bias from perceptual effects. *Q J Exp Psychol (Hove)* 66: 319-337, 2012.

Garcia-Perez MA, and Alcala-Quintana R. Sampling plans for fitting the psychometric function. *The Spanish journal of psychology* 8: 256-289, 2005.

Garcia-Perez MA, and Peli E. The bisection point across variants of the task. *Atten Percept Psychophys* 76: 1671-1697, 2014.

Goris RL, Movshon JA, and Simoncelli EP. Partitioning neuronal variability. *Nature neuroscience* 17: 858-865, 2014.

Green D, and Swets J. *Signal detection theory and psychophysics*. New York: John Wiley and Sons, Inc., 1966.

Green DM. Stimulus selection in adaptive psychophysical procedures. *J Acoust Soc Am* 87: 2662-2674, 1990.

Grimaldi P, Lau H, and Basso MA. There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience and biobehavioral reviews* 55: 88-97, 2015.

Hall J. Maximum ‐ Likelihood Sequential Procedure for Estimation of Psychometric Functions. *The Journal of the Acoustical Society of America* 44: 370-370, 1968.

Hall JL. Hybrid adaptive procedure for estimation of psychometric functions. *J Acoust Soc Am* 69: 1763-1769, 1981.

Harvey LO. Efficient estimation of sensory thresholds. *Behavior Research Methods, Instruments, & Computers* 18: 623-632, 1986.

Hsu YF, and Doble CW. A threshold theory account of psychometric functions with response confidence under the balance condition. *The British journal of mathematical and statistical psychology* 68: 158-177, 2015.

Juslin P, Olsson H, and Winman A. The Calibration Issue: Theoretical Comments on Suantak, Bolger, and Ferrell (1996). *Organizational behavior and human decision processes* 73: 3-26, 1998.

Kaernbach C. Simple adaptive testing with the weighted up-down method. *Perception & psychophysics* 49: 227-229, 1991.

Karmali F, Chaudhuri SE, Yi Y, and Merfeld DM. Determining thresholds using adaptive procedures and psychometric fits: evaluating efficiency using theory, simulations, and human experiments. *Experimental brain research* 1-17, 2015.

Keren G. Calibration and probability judgements: Conceptual and methodological issues. *Acta psychologica* 77: 217-273, 1991.

Kontsevich LL, and Tyler CW. Bayesian adaptive estimation of psychometric slope and threshold. *Vision Res* 39: 2729-2737, 1999.

Lau H, and Rosenthal D. Empirical support for higher-order theories of conscious awareness. *Trends in cognitive sciences* 15: 365-373, 2011.

Leek MR. Adaptive procedures in psychophysical research. *Perception & psychophysics* 63: 1279-1292, 2001.

Lichtenstein S, and Fischhoff B. Training for calibration. *Organizational Behavior and Human Performance* 26: 149-171, 1980.

Lichtenstein S, Fischhoff B, and Phillips L. Calibration of probabilities: The state of the art to 1980. . In: *Judgement under uncertainty: Heuristics and biases*, edited by Kahneman D, Slovic P, and Tverski A. New York: Cambridge University Press, 1982.

Lim K, and Merfeld DM. Signal detection theory and vestibular perception: II. Fitting perceptual thresholds as a function of frequency. *Exp Brain Res* 222: 303-320, 2012.

Linschoten MR, Harvey LO, Jr., Eller PM, and Jafek BW. Fast and accurate measurement of taste and smell thresholds using a maximum-likelihood adaptive staircase procedure. *Perception & psychophysics* 63: 1330-1347, 2001.

Macmillan NA, and Creelman CD. *Detection Theory: A User's Guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2005.

Merfeld DM. Signal detection theory and vestibular thresholds: I. Basic theory and practical considerations. *Exp Brain Res* 210: 389-405, 2011.

Merfeld DM, Clark TK, Lu YM, and Karmali F. Dynamics of Individual Perceptual Decisions. *Journal of neurophysiology* jn. 00225.02015, 2015.

Merfeld DM, Priesol A, Lee D, and Lewis RF. Potential solutions to several vestibular challenges facing clinicians. *Journal of vestibular research : equilibrium & orientation* 20: 71-77, 2010.

Murphy AH. A new vector partition of the probability score. *Journal of Applied Meteorology* 12: 595-600, 1973.

Murphy AH. Scalar and vector partitions of the probability score: Part I. Two-state situation. *Journal of Applied Meteorology* 11: 273-282, 1972.

Okamoto Y. An experimental analysis of psychometric functions in a threshold discrimination task with four response categories. *Japanese Psychological Research* 54: 368-377, 2012.

Pentland A. Maximum likelihood estimation: the best PEST. *Perception & psychophysics* 28: 377-379, 1980.

Sawides L, Dorronsoro C, Haun AM, Peli E, and Marcos S. Using pattern classification to measure adaptation to the orientation of high order aberrations. *PloS one* 8: e70856, 2013.

Shen Y, Dai W, and Richards VM. A MATLAB toolbox for the efficient estimation of the psychometric function using the updated maximum-likelihood adaptive procedure. *Behavior research methods* 47: 13-26, 2015.

Shen Y, and Richards VM. A maximum-likelihood procedure for estimating psychometric functions: thresholds, slopes, and lapses of attention. *The Journal of the Acoustical Society of America* 132: 957-967, 2012.

Stankov L. Calibration curves, scatterplots and the distinction between general knowledge and perceptual tasks. *Learning and Individual Differences* 10: 29-50, 1998.

Stankov L, Pallier G, Danthiir V, and Morony S. Perceptual Underconfidence: a conceptual illusion? *European Journal of Psychological Assessment* 28: 190-200, 2012.

Stone ER, and Opel RB. Training to improve calibration and discrimination: The effects of performance and environmental feedback. *Organizational behavior and human decision processes* 83: 282-309, 2000.

Suantak L, Bolger F, and Ferrell WR. The hard–easy effect in subjective probability calibration. *Organizational behavior and human decision processes* 67: 201-221, 1996.

Taylor MM, and Creelman CD. PEST: Efficient estimates on probability functions. *J Acoust Soc Am* 41: 782-787, 1967.

Tolhurst D, Movshon J, and Thompson I. The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental brain research* 41: 414-419, 1981.

Treutwein B. Adaptive psychophysical procedures. *Vision Res* 35: 2503-2522, 1995.

Valko Y, Lewis RF, Priesol AJ, and Merfeld DM. Vestibular labyrinth contributions to human whole-body motion discrimination. *J Neurosci* 32: 13537-13542, 2012.

Watson AB, and Pelli DG. QUEST: a Bayesian adaptive psychometric method. *Perception & psychophysics* 33: 113-120, 1983.

Watt R, and Andrews D. APE: Adaptive probit estimation of psychometric functions. *Current Psychological Reviews* 1: 205-213, 1981.

Wetherill GB. Sequential estimation of quantal response curves. *Journal of the Royal Statistics Society* B25: 1-48, 1963.

Wetherill GB, and Levitt H. Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology* 18: 1-10, 1965.

Wichmann FA, and Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & psychophysics* 63: 1293-1313, 2001a.

Wichmann FA, and Hill NJ. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & psychophysics* 63: 1314-1329, 2001b.

Yates JF. *Judgment and decision making*. Prentice-Hall, Inc, 1990.

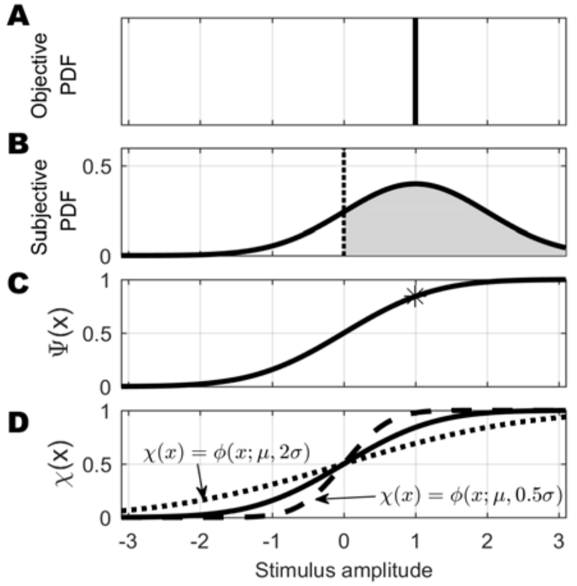


Fig. 1. Relationship between decision variables, psychometric functions ($\Psi(x)$), and confidence functions ($\chi(x)$) in our confidence signal detection (CSD) model. *A:* The stimulus for this example is well controlled having an amplitude of +1.0 with little variation, so the objective probability density function (PDF) is a delta function. *B:* A signal detection model assumes additive noise. For this example, Gaussian noise having zero-mean and a standard deviation of 1 is added to the stimulus of +1.0 and leads to the subjective PDF shown. The dotted vertical line at zero represents a decision boundary. If a sampled decision variable falls to the right of the decision boundary, represented by the gray area, the subject decides positive. If the sampled decision variable falls to the left, the subject decides negative. For this example, 84% of the decision variables lead to the subject deciding positive. *C:* The asterisk, located at (1, 0.84) represents the example data point illustrated in the previous panel. When this process is repeated for a variety of different stimulus levels, it yields a psychometric function, $\Psi(x)$ (black curve). *D:* Similarly, a relationship between confidence and the stimulus for an individual trial can be represented by a confidence function ($\chi(x)$). The psychometric function represents average subject performance, and confidence is defined as well-calibrated when confidence matches average subject performance (Bjorkman et al. 1993; Ferrell 1995; Keren 1991; Lichtenstein et al. 1982; Stankov et al. 2012). Therefore, well-calibrated confidence matches the psychometric function ($\chi(x) = \Psi(x)$) and is plotted as the solid curve. Also shown are confidence functions that represent over-confidence (dashed) and under-confidence (dotted).

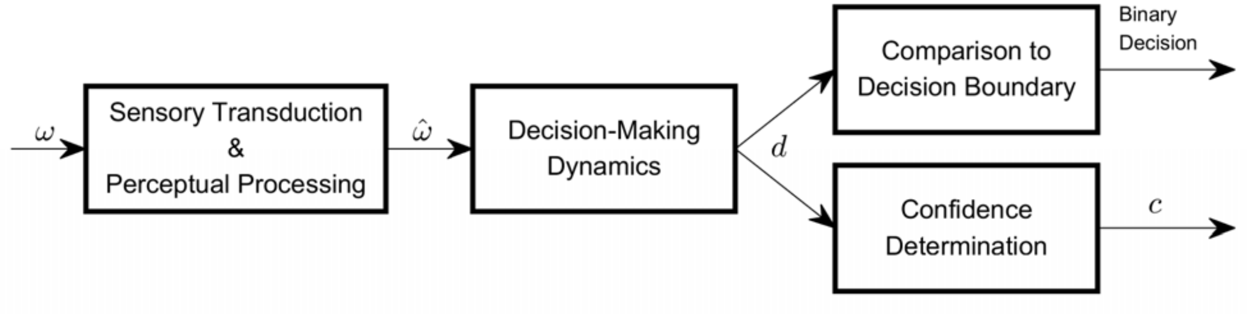


Fig. 2. Schematic representation of pertinent neural processing. For the experimental investigations herein, the stimulus is angular velocity (ω), which is transduced and processed via pertinent perceptual mechanism to yield a perceptual representation of angular velocity ($\hat{\omega}$); these representations could readily be generalized to any perceptual process. This perceptual signal may be further filtered (e.g. Merfeld et al. 2015) to yield a decision-variable (*d*) used both to make the binary decision – by comparing the decision variable to the decision boundary – and via additional neural manipulations to yield confidence (*c*) in the decision. The latter confidence calculations may be performed "correctly" [i.e., $c = \chi(d) = \phi(d, \mu, \sigma)$], yielding well-calibrated data. Alternatively, these confidence calculations could be miscalibrated [i.e., $c = \chi(d) = \phi(d, \mu, k\sigma)$], where $k \neq 1$, leading to overconfidence ($k < 1$) or under-confidence ($k > 1$). Furthermore, as we will evaluate via simulations, the confidence function may be independent of the noise distribution that defines the psychometric function (i.e., confidence function is not Gaussian for this application).

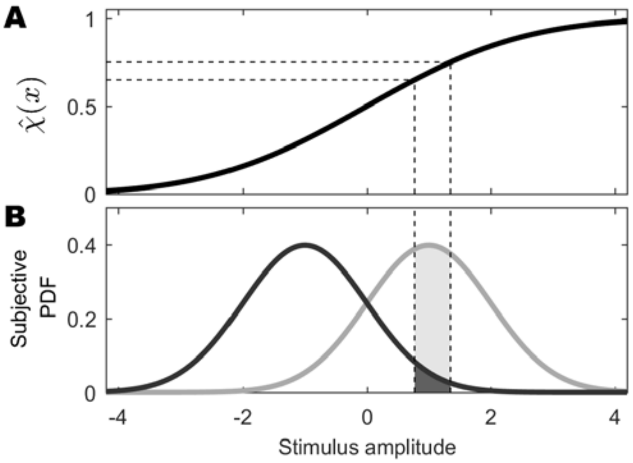


Fig. 3. Illustration of how confidence probability judgments from individual trials contribute to a maximum likelihood psychometric function fit. *A:* Given a confidence probability judgment, we can use the inverse fitted confidence function ($\hat{\chi}^{-1}(c_j)$) to calculate a modeled decision variable to accompany that judgment. More specifically, given upper and lower limits to a confidence probability judgment (dashed horizontal lines), we can use the inverse fitted confidence function to calculate the corresponding upper and lower decision variable limits (dashed vertical lines). *B:* Given the estimated decision variable range shown by the dashed vertical lines, we can calculate the probability that the given stimulus (s_j) and psychometric function noise model would yield that confidence probability judgment. Two examples are illustrated. The light curve shows the decision variable PDF for the stimulus having an amplitude of +1.0 shown in Fig. 1B and the light shaded area represents the probability of the confidence probability judgment for a +1.0 stimulus. The dark curve shows the PDF for a stimulus having an amplitude of -1.0, and the dark shaded area represents the probability for a -1.0 stimulus. High confidence that the motion is positive is much more probable (i.e., much more likely) for the +1 stimulus than for the -1 stimulus.

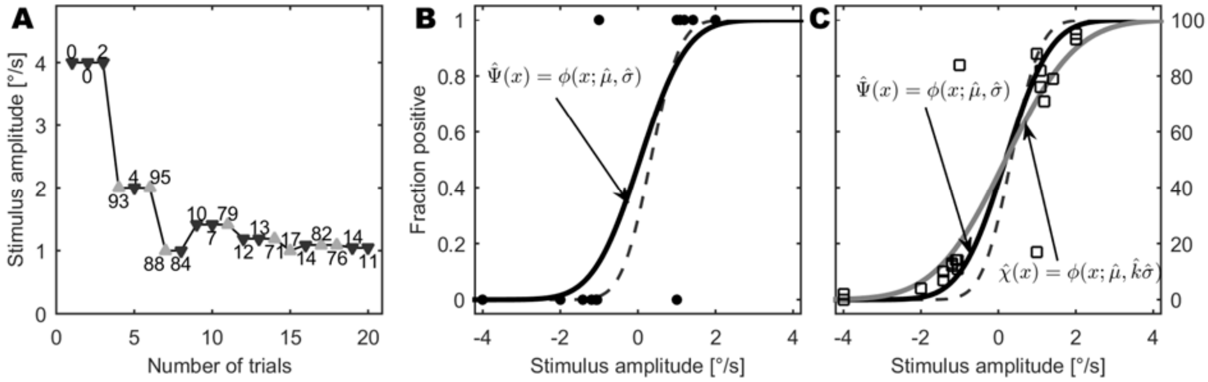


Fig. 4. Example fits for a human test. *A:* Example stimulus track, including confidence probability judgments, is shown for first 20 trials. Upward-pointing gray triangles and downward-pointing black triangles represent rightward and leftward trials, respectively. *B:* Following 20 binary forced-choice trials, a conventional psychometric function (black curve), $\hat{\Psi}(x) = \phi(x; \hat{\mu} = 0.05, \hat{\sigma} = 0.95)$, was fit to the binary forced-choice data points shown. *C:* Given the same 20 trials with confidence probability judgments, a psychometric function (black curve), $\hat{\Psi}(x) = \phi(x; \hat{\mu} = 0.19, \hat{\sigma} = 0.91)$, and a confidence function (gray curve), $\hat{\chi}(x) = \phi(x; \hat{\mu} = 0.19, \hat{k}\hat{\sigma} = 1.43)$, were simultaneously fit to the confidence data. All example data are from one of the human data sets (Fig. 5, 4th column) presented herein. For comparison, the fitted psychometric function determined after 100 binary forced-choice trials using conventional methods, $\hat{\Psi}(x) = \phi(x; \hat{\mu} = 0.33, \hat{\sigma} = 0.59)$, is also shown via dashed lines on panels b and c. Half-scale (50% to 100%) probability judgments provided by subjects have been converted to full-scale (0 to 100%) judgments as described in Methods.

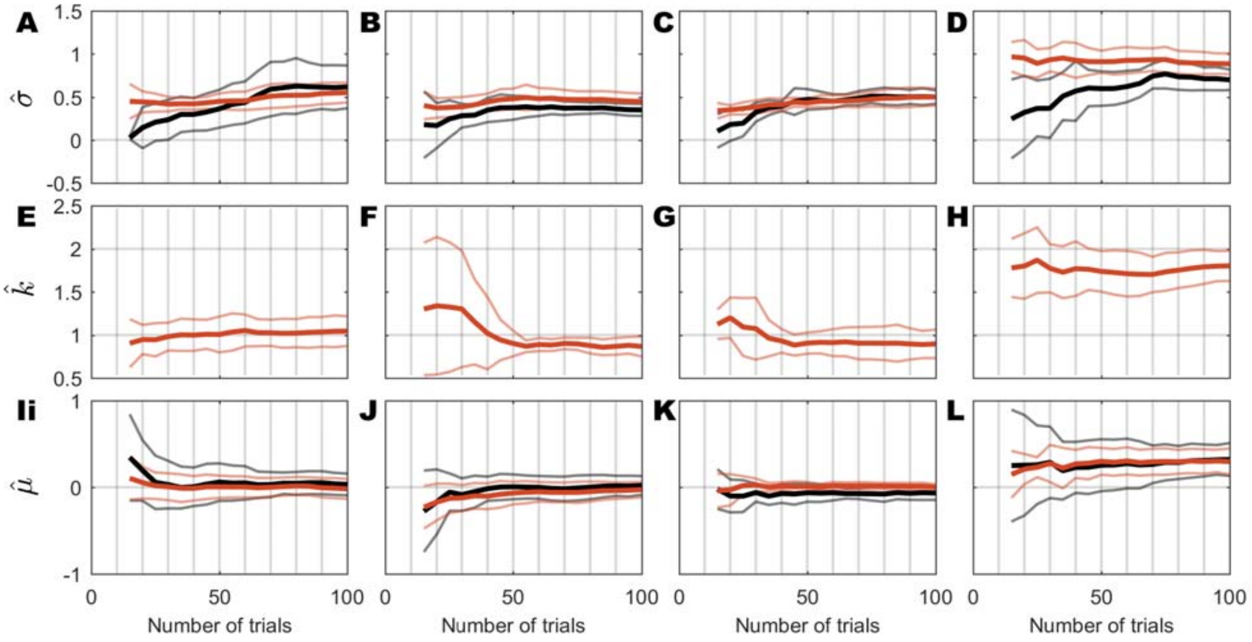


Fig. 5. Summary of human psychometric parameter estimates as trial number increases. Each column represents fitted parameters for one subject. Top row (A-D) shows average fitted psychometric width parameter ($\hat{\sigma}$). Middle row (E-H) shows average fitted confidence scaling factor ($\hat{k}$). Bottom row (I-L) shows average fitted psychometric function bias ($\hat{\mu}$). Thick black curves show average psychometric parameter estimates calculated using conventional forced-choice analyses. Thick red curves show average parameter estimates determined by fitting confidence probability judgment data. Errors bars (thin gray curves and thin red curves, respectively) represent standard deviation of parameter estimates.

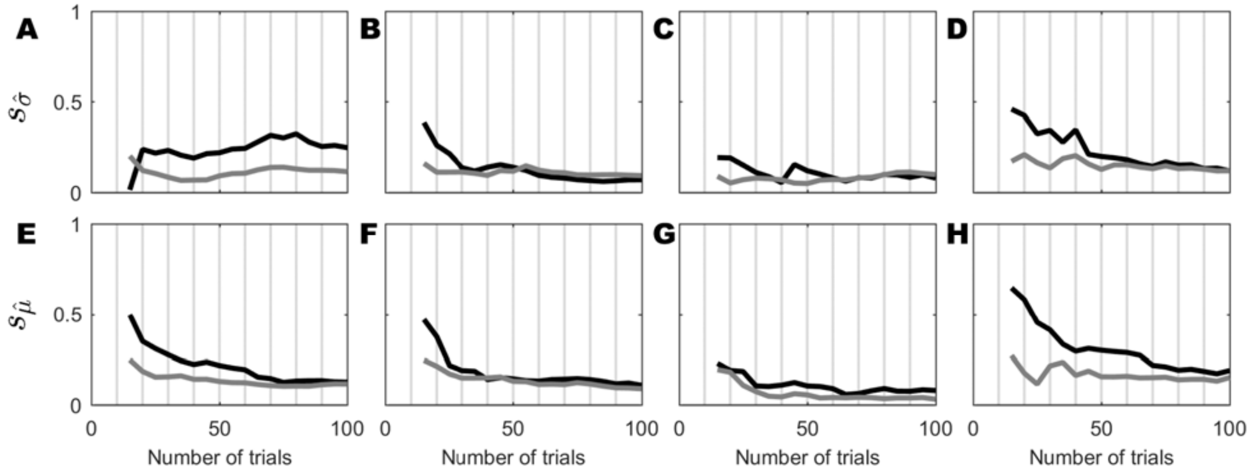


Fig. 6. Standard deviation of human psychometric parameter estimates as trial number increases. Each column represents fitted parameters for one subject in the same order as Fig. 5. Top row (A-D) represents the standard deviation of the fitted psychometric width parameter ($\hat{\sigma}$). Bottom row (E-H) represents the fitted psychometric function bias ($\hat{\mu}$). Black curves show standard deviation of psychometric parameter estimates calculated using conventional forced-choice analyses. Gray curves show standard deviation of parameter estimates determined via our CSD model fit.

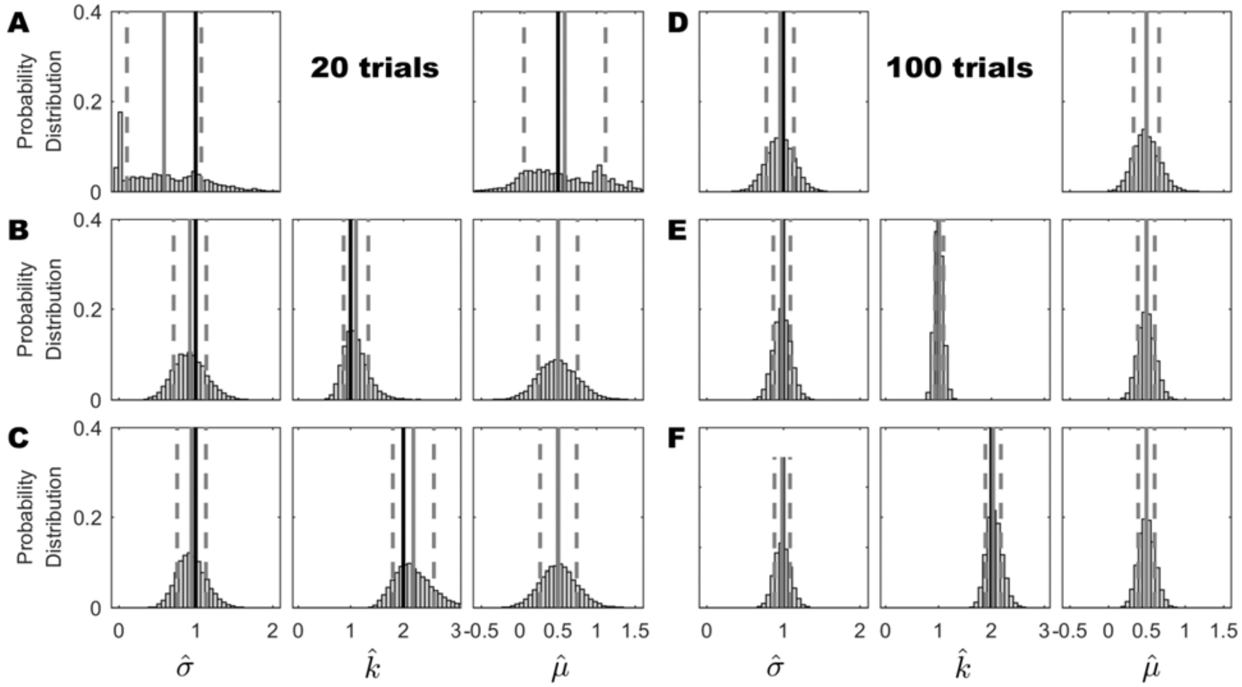


Fig 7. Parameter distributions show parameter estimates for 10,000 simulated experiments with 20 and 100 trials. The columns from left to right represent the fitted psychometric width parameter ($\hat{\sigma}$), the fitted confidence scaling factor ($\hat{k}$) and the fitted psychometric function bias ($\hat{\mu}$) as shown on the x-axis at bottom. Top row (A and D) represents fitted parameters of conventional binary forced-choice parameter estimates. Middle row (B and E) represents fitted parameters estimates determined via our CSD model fit for a well-calibrated subject ($k = 1$). Bottom row (C and F) represents fitted parameters estimates determined via our CSD model fit for an under-confident subject ($k = 2$). The solid black line shows the actual parameter value (i.e. $\mu = 0.5$ or $\sigma = 1$), the solid gray line shows the mean of fitted parameters, and the dashed gray lines indicate standard deviation on each side of the mean.

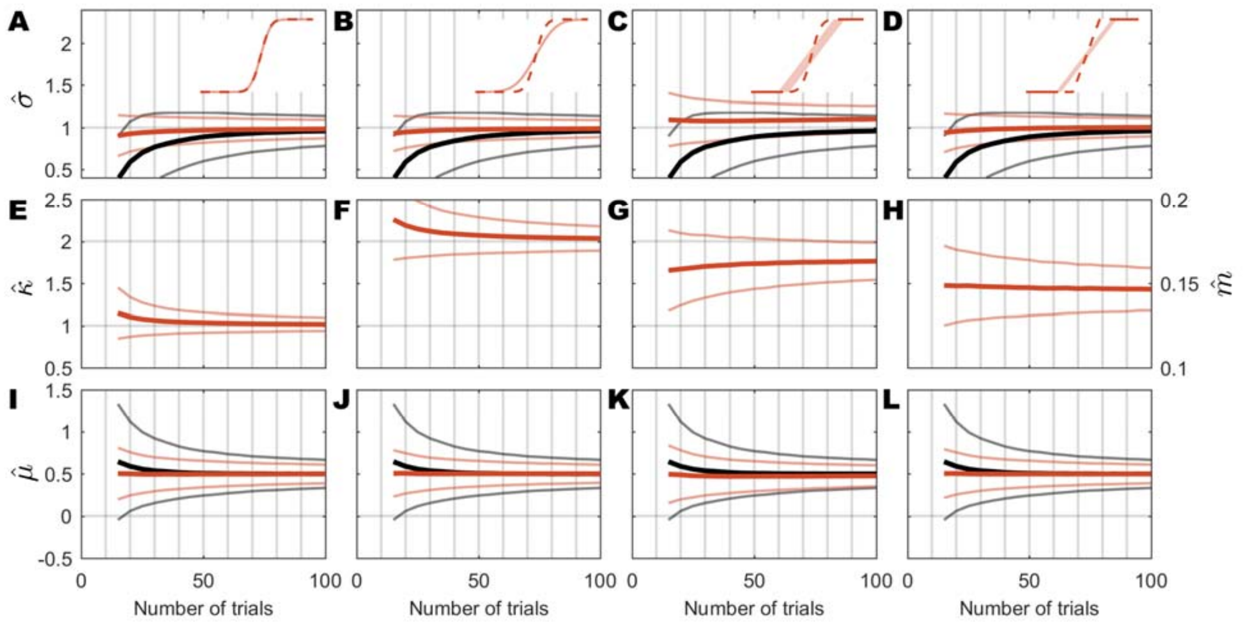


Fig. 8. Summary of simulation parameter estimates as trial number increases. As illustrated via insets, each column represents different simulated combinations of the confidence function (red solid curves) and the fitted confidence function (red dashed curves). First column shows a well-calibrated subject ($k=1$) when both confidence and confidence fit functions are cumulative Gaussians. Second column shows an under-confident subject ($k=2$) when both confidence and confidence fit functions are cumulative Gaussians. Third column shows an under-confident subject when the confidence function is linear, $\chi(x)=m(x-\mu)+0.5=0.1443x+0.428$, with added zero-mean uniform noise ($U(-0.1,+0.1)$), and the confidence fit function is a cumulative Gaussian. Fourth column shows an under-confident subject with the same linear confidence function with added zero-mean uniform noise ($U(-0.05,+0.05)$) when the confidence fit function is linear, $\hat{\chi}(x)=\hat{m}(x-\hat{\mu})+0.5$. Top row (A-D) shows fitted psychometric width parameter ($\hat{\sigma}$). Middle row (E-G) shows fitted confidence-scaling factor ($\hat{k}$) or (H) fitted slope of confidence function. Bottom row (I-L) shows fitted psychometric function bias ($\hat{\mu}$). Thick black curves show average conventional forced-choice parameter estimates, which are identical for all conditions. Thick red curves show average parameter estimates determined by fitting confidence probability judgments. Errors bars (thin gray curves and thin red curves, respectively) represent standard deviation of parameter estimates.

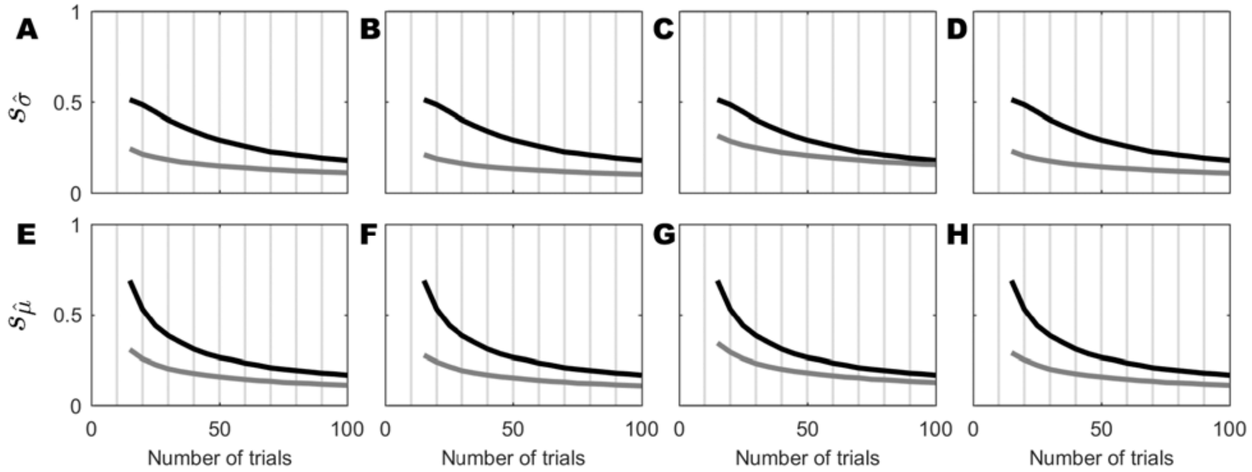


Fig. 9. Standard deviation of simulation parameter estimates as trial number increases. Each column represent the same conditions as Fig. 8. Top row (A-D) represents the fitted psychometric width parameter ($\hat{\sigma}$). Bottom row (E-G) represents the fitted psychometric function bias ($\hat{\mu}$). Black curves show standard deviation of conventional forced-choice parameter estimates, which are identical for all conditions. Gray curves show standard deviation of parameter estimates determined via our CSD model fit.

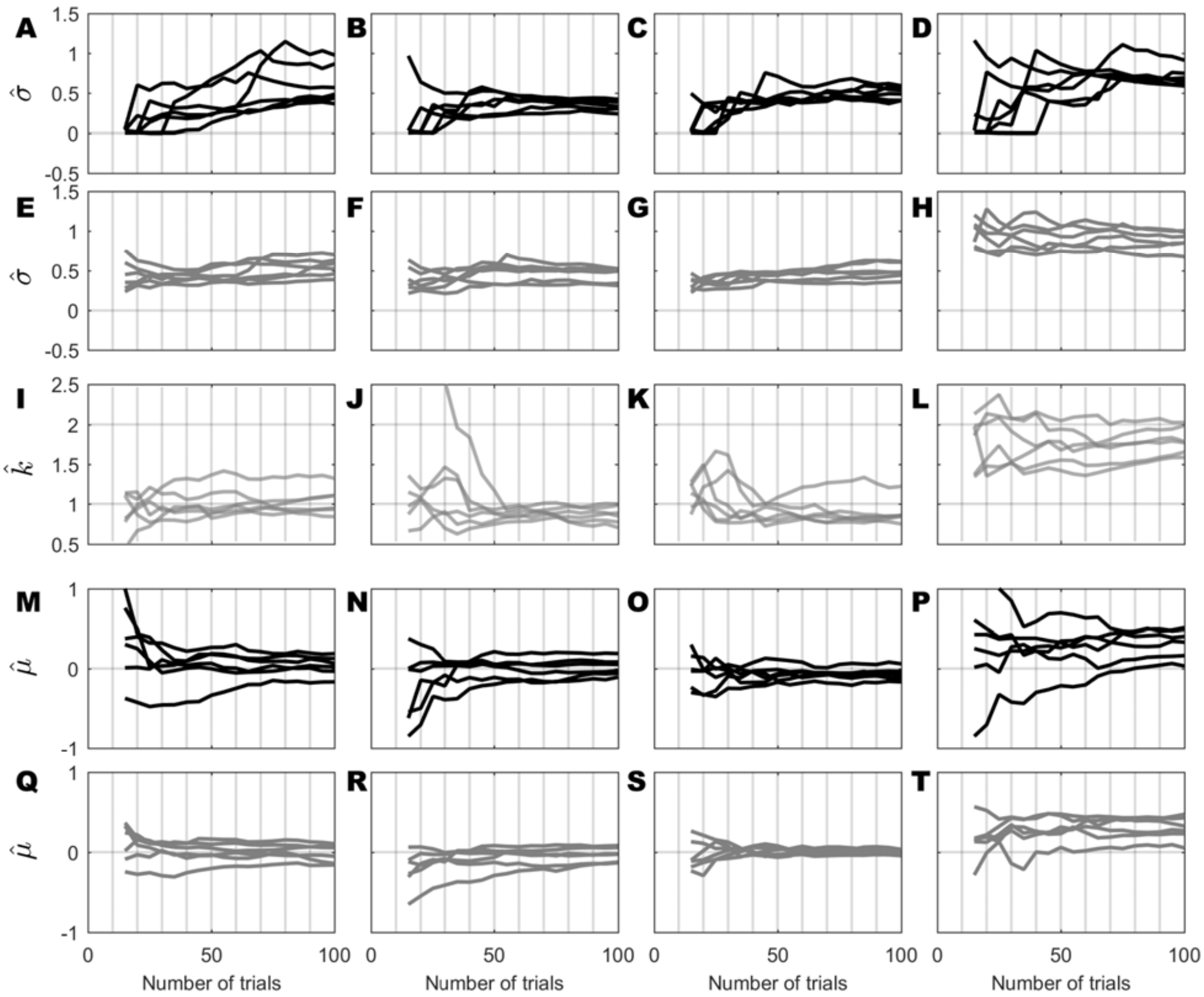


Fig. B1. Human psychometric width parameter ($\hat{\sigma}$), confidence scaling factor ($\hat{k}$) and, bias parameter ($\hat{\mu}$) estimates as trial number increases for each subject for each of 6 test sessions. Each column shows fitted parameters for one subject. Top row (A-D) shows fitted psychometric width parameter using conventional forced-choice analyses. Second row (E-H**)** shows fitted psychometric width parameter for CSD model fit. Third row (I-L) shows fitted confidence scaling factor for CSD model fit. Fourth row (M-P) shows fitted psychometric function bias using conventional forced-choice analyses. Bottom row (Q-T) shows fitted psychometric function bias for CSD model fit.

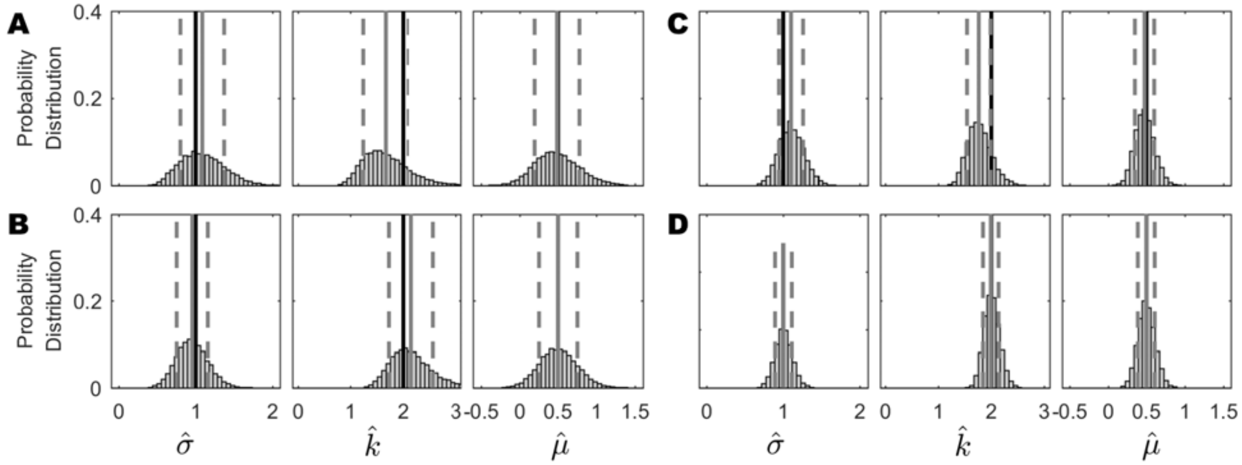
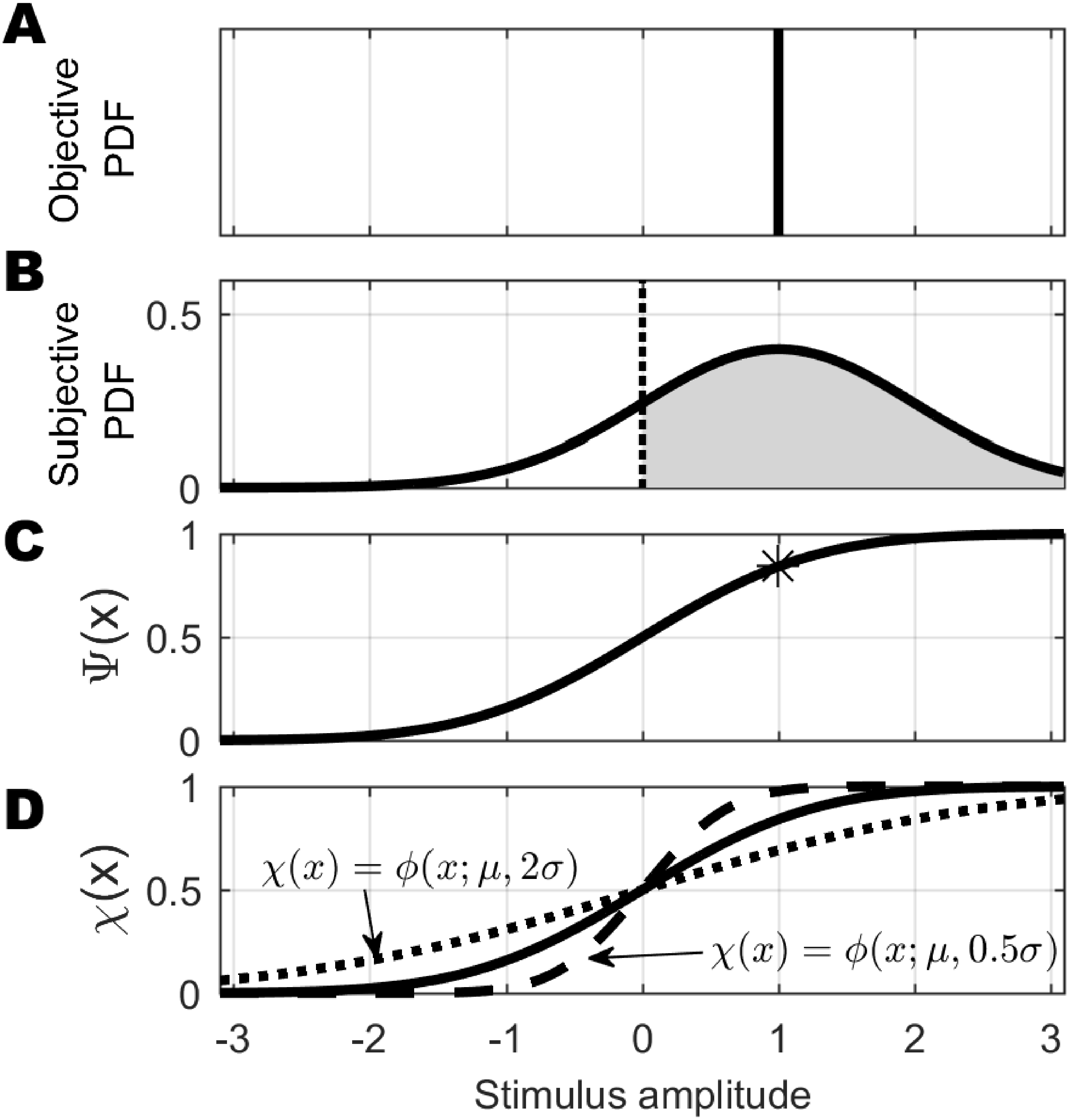
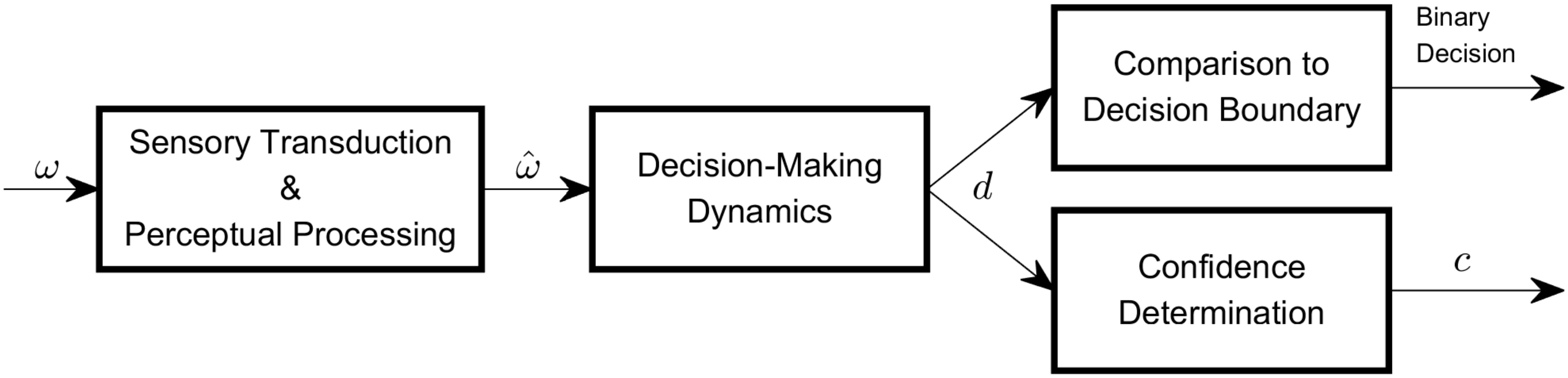
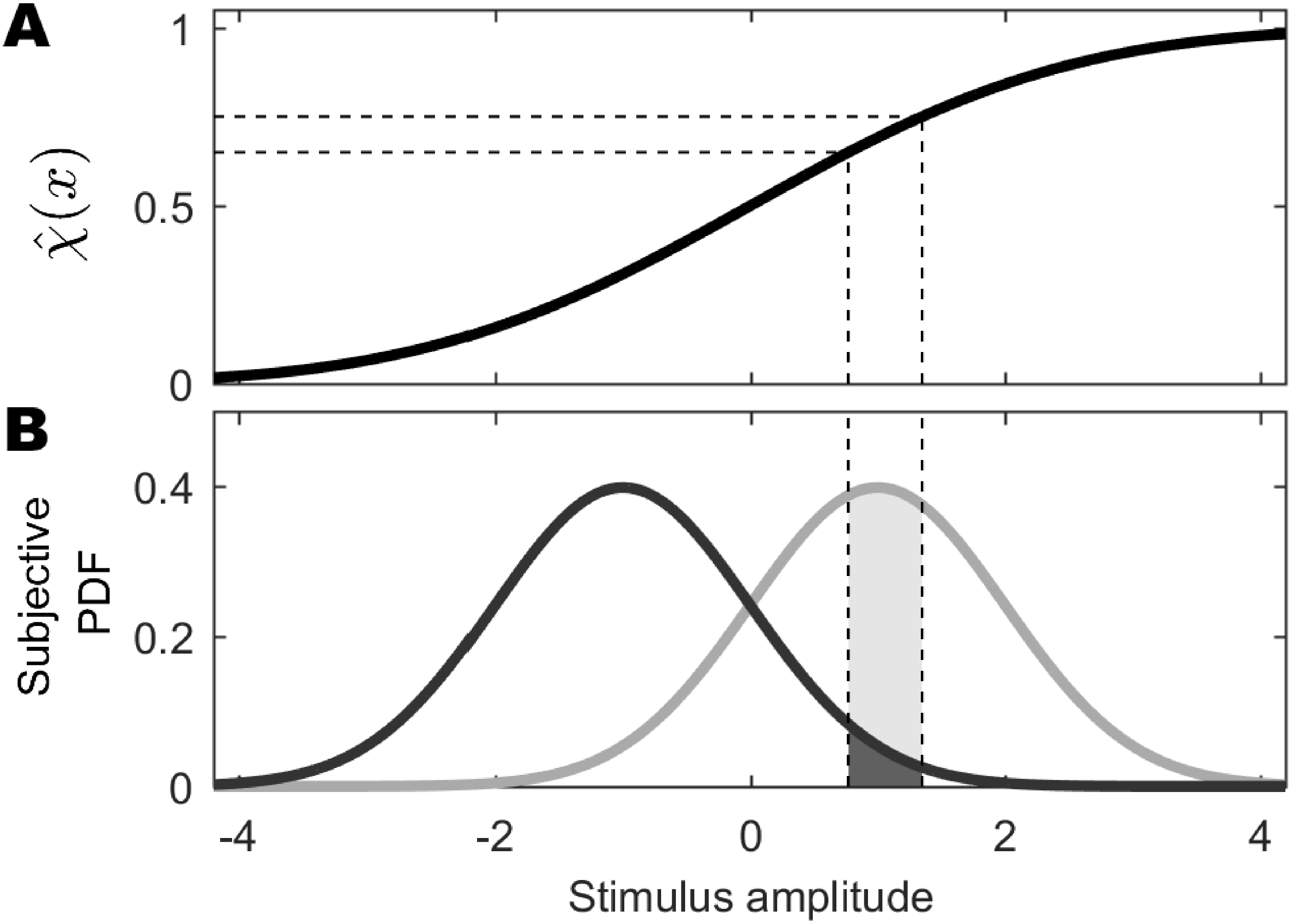
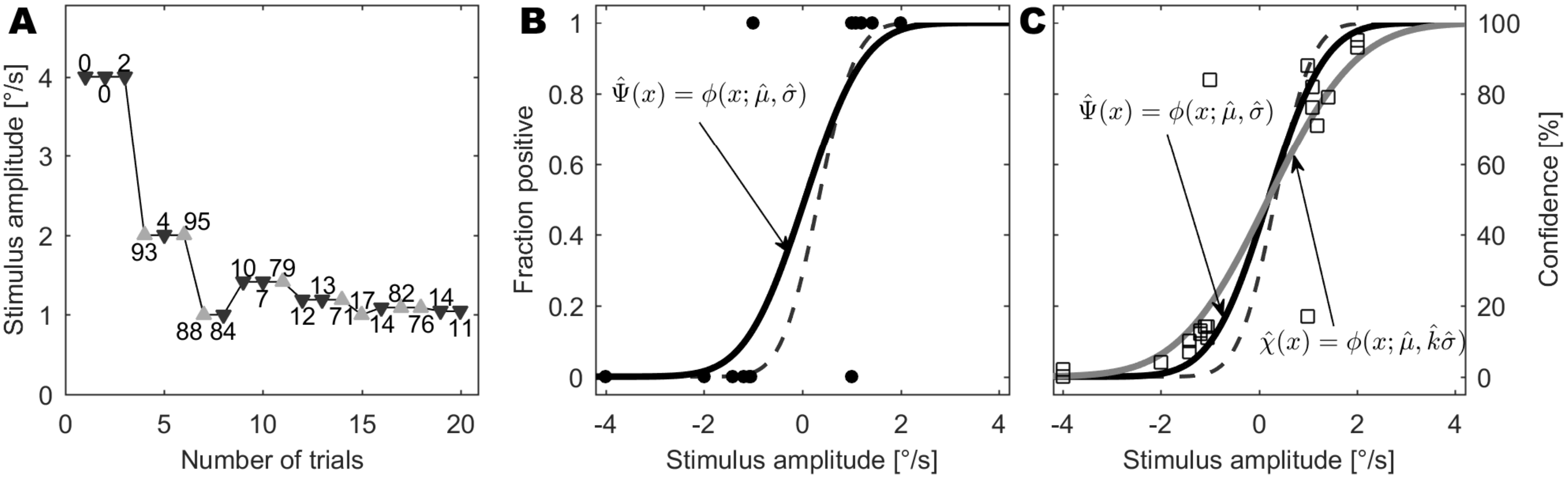


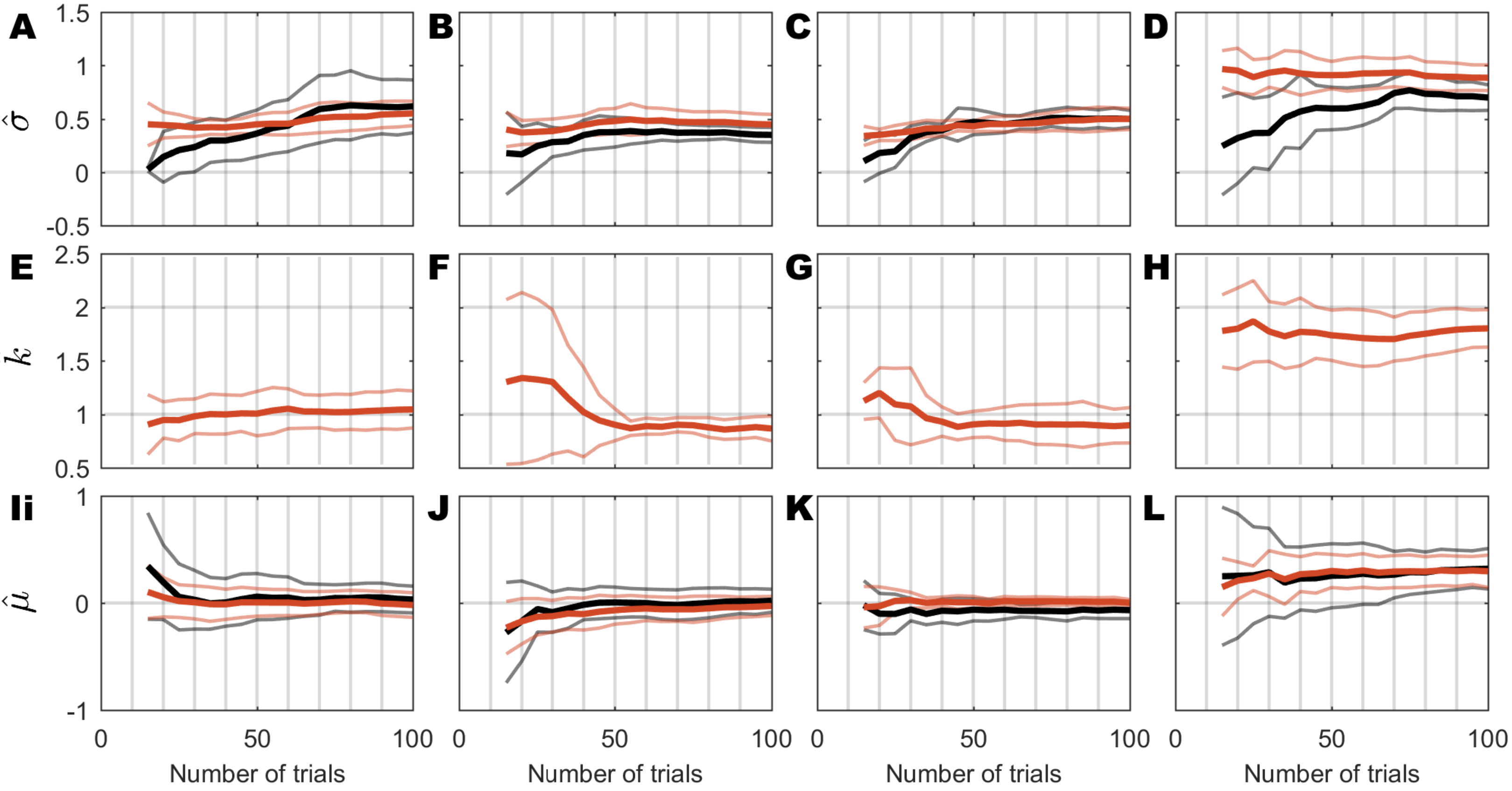
Fig. B2. Parameter distributions show parameter estimates for 10,000 simulated experiments with 20 and 100 trials in the same order as figure 6. Top row (A and C) represents fitted parameters estimates determined via our CSD model fit for an under-confident subject when the confidence function is linear, $\chi(x) = m(x-\mu)+0.5 = 0.1443x+0.428$, with added zero-mean uniform noise ($U(-0.1,+0.1)$), and the confidence fit function is a cumulative Gaussian. Bottom row (B and D) represents fitted parameters estimates determined via our CSD model fit for an under-confident subject with the same linear confidence function with added zero-mean uniform noise ($U(-0.05,+0.05)$) when the confidence fit function is linear, $\hat{\chi}(x) = \hat{m}(x-\hat{\mu})+0.5$. The solid black line shows the actual parameter value (i.e. $\mu = 0.5$ or $\sigma = 1$), the solid gray line shows the mean of fitted parameters, and the dashed gray lines indicate standard deviation on each side of the mean.

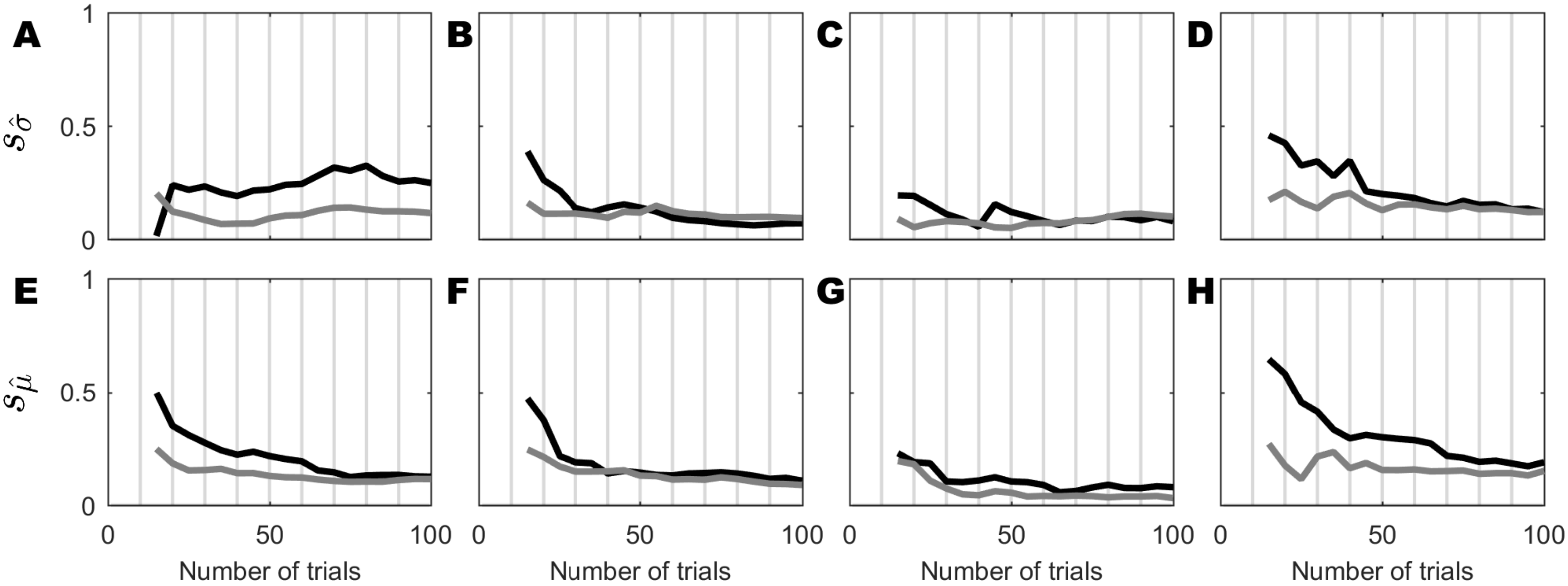

A
Objective PDF
B
Subjective PDF
0.5
0
C
Ψ(x)
1
0.5
0
D
χ(x)
1
0.5
0
$\chi(x) = \phi(x; \mu, 2\sigma)$
$\chi(x) = \phi(x; \mu, 0.5\sigma)$
-3
-2
-1
0
1
2
3
Stimulus amplitude

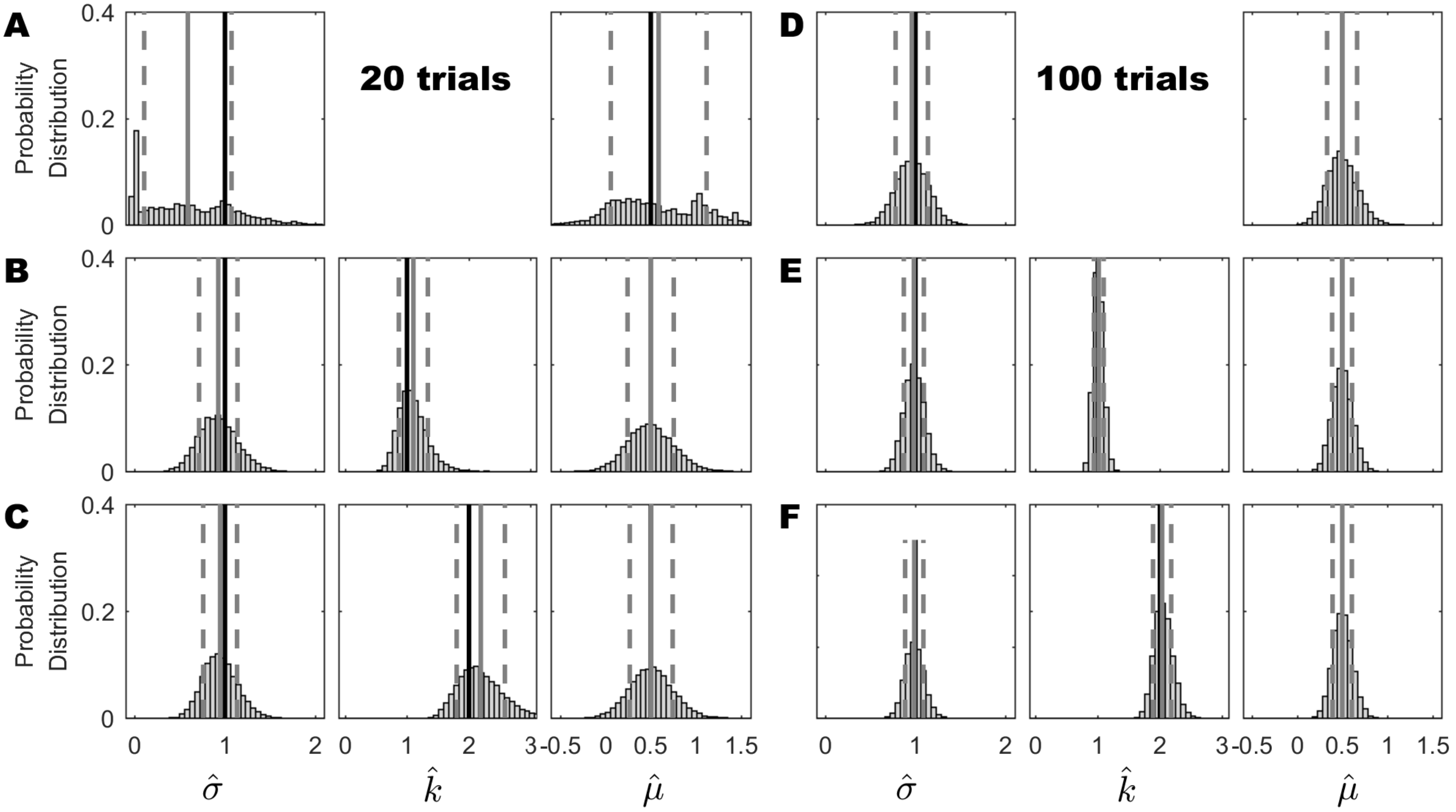

ω
Sensory Transduction
&
Perceptual Processing
$\hat{\omega}$
Decision-Making
Dynamics
d
Comparison to
Decision Boundary
Binary
Decision
Confidence
Determination
c

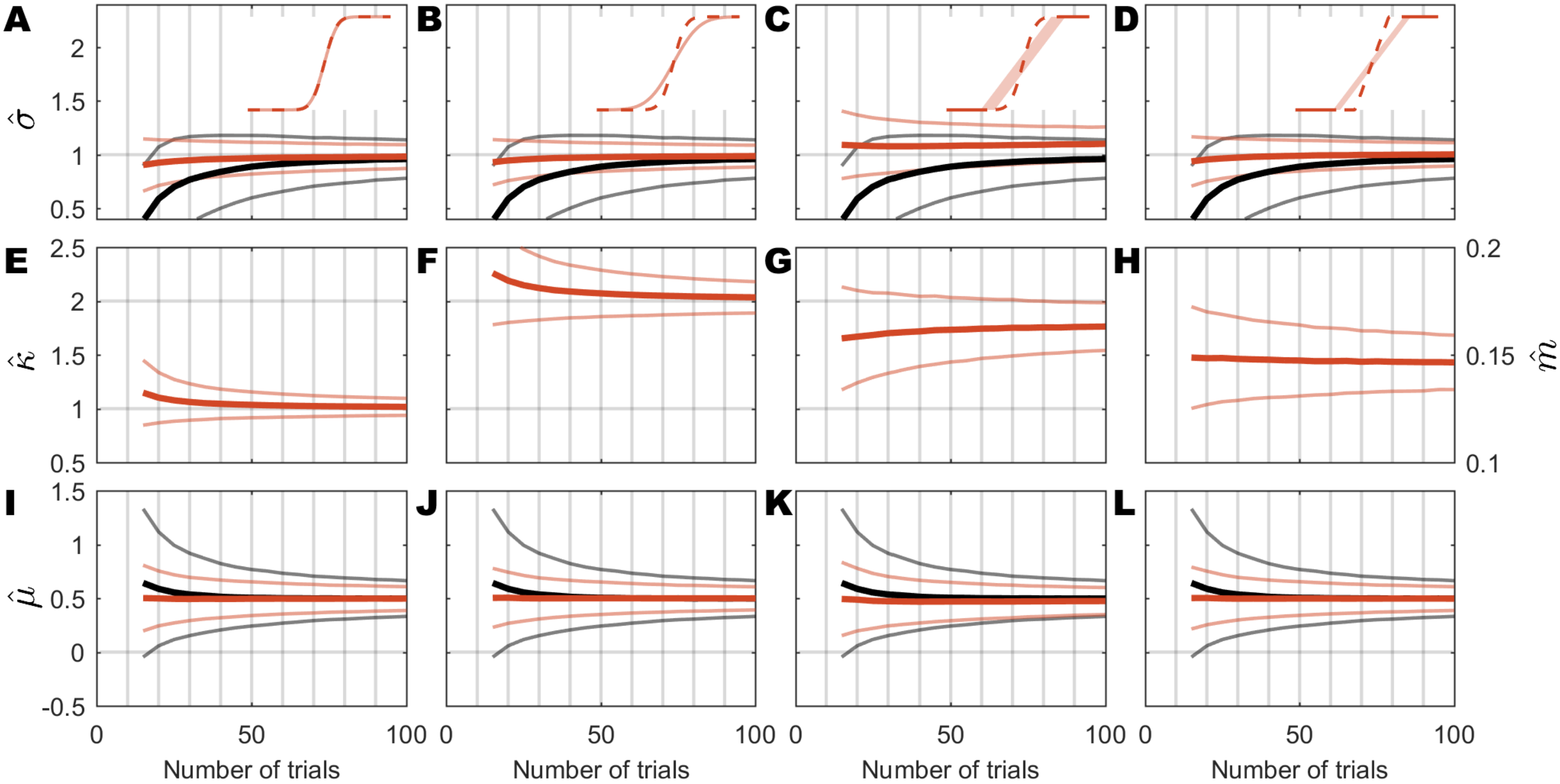

A
B
1
0.5
0
$\hat{\chi}(x)$
0.4
0.2
0
Subjective PDF
-4
-2
0
2
4
Stimulus amplitude

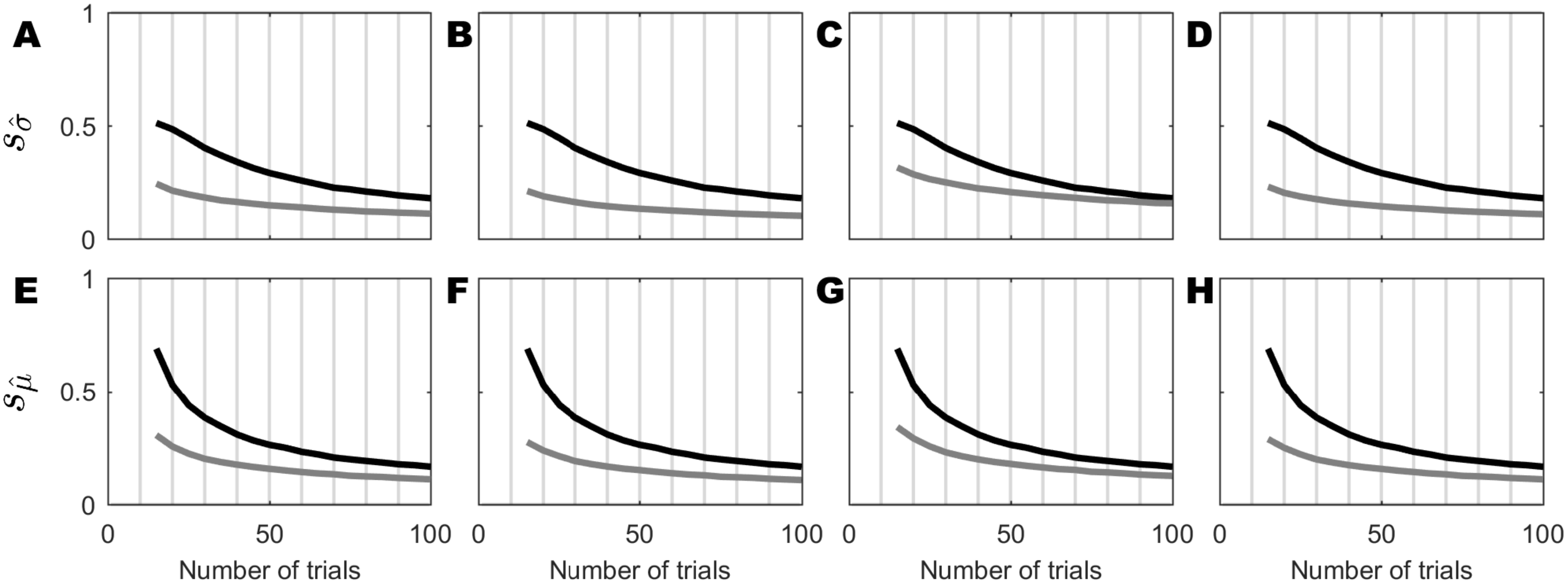

A
Stimulus amplitude [°/s]
Number of trials
0
1
2
3
4
0
5
10
15
20
0
0
2
93
4
95
88
84
10
7
79
12
13
71
17
14
82
76
14
11
B
Fraction positive
Stimulus amplitude [°/s]
0
0.2
0.4
0.6
0.8
1
-4
-2
0
2
4
$\hat{\Psi}(x) = \phi(x; \hat{\mu}, \hat{\sigma})$
C
Confidence [%]
Stimulus amplitude [°/s]
0
20
40
60
80
100
-4
-2
0
2
4
$\hat{\Psi}(x) = \phi(x; \hat{\mu}, \hat{\sigma})$
$\hat{\chi}(x) = \phi(x; \hat{\mu}, \hat{k}\hat{\sigma})$

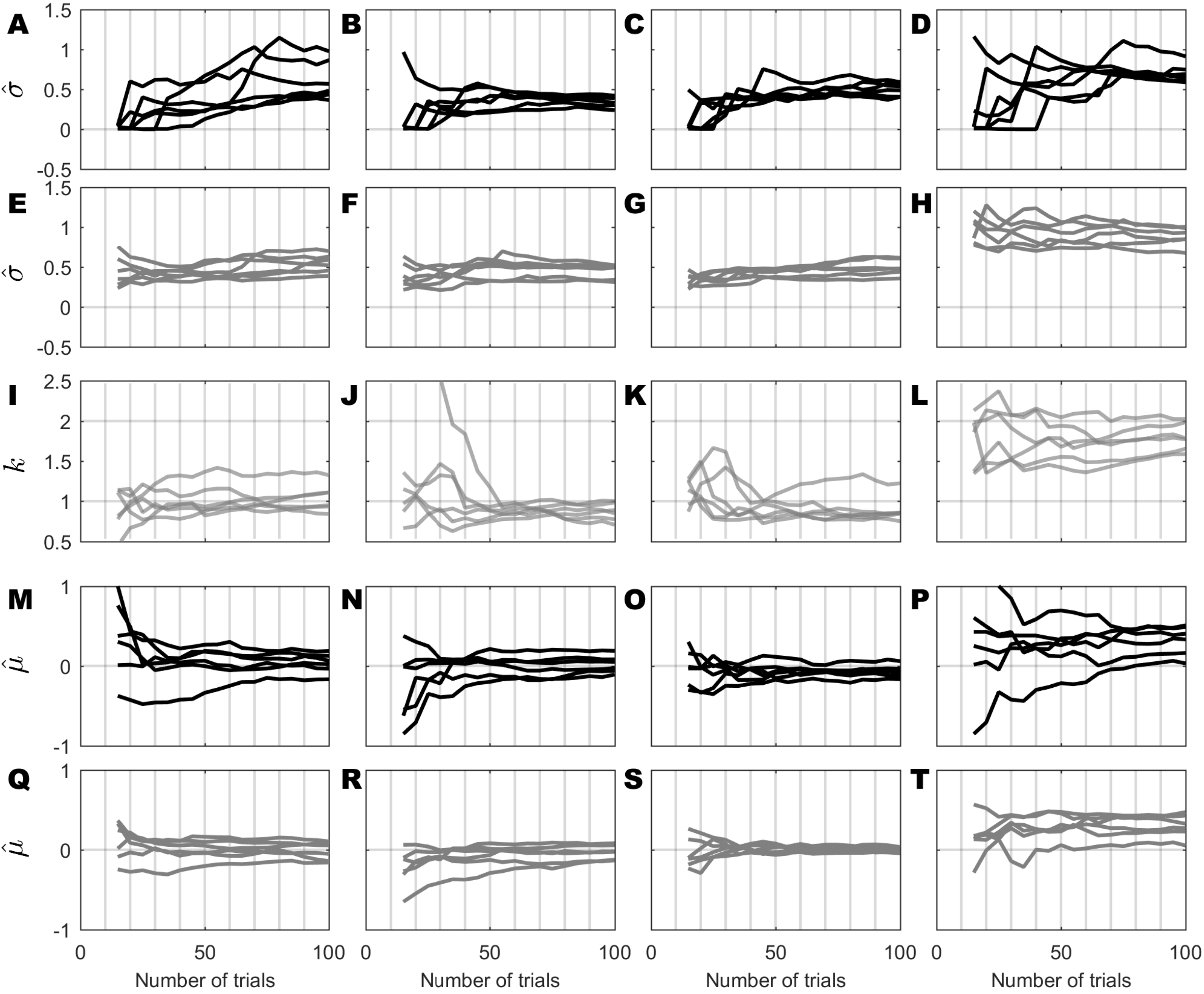

A
B
C
D
E
F
G
H
Ii
J
K
L
$\hat{\sigma}$
$\hat{k}$
$\hat{\mu}$
1.5
1
0.5
0
-0.5
2.5
2
1.5
1
0.5
1
0
-1
0
50
100
Number of trials

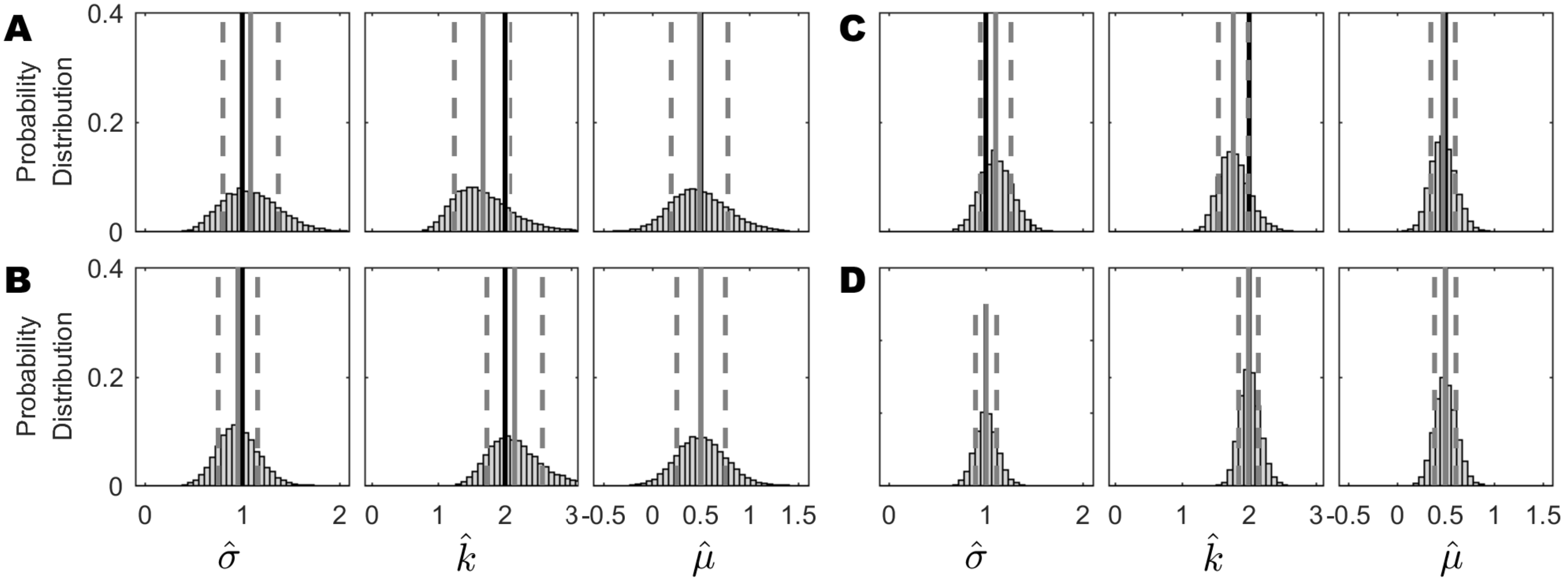

A
B
C
D
E
F
G
H
$s_{\hat{\sigma}}$
$s_{\hat{\mu}}$
1
0.5
0
1
0.5
0
0
50
100
Number of trials
0
50
100
Number of trials
0
50
100
Number of trials
0
50
100
Number of trials


A
B
C
D
E
F
20 trials
100 trials
Probability Distribution
0.4
0.2
0
0
1
2
3
-0.5
0.5
1.5
$\hat{\sigma}$
$\hat{k}$
$\hat{\mu}$


A
B
C
D
E
F
G
H
I
J
K
L
$\hat{\sigma}$
$\hat{\kappa}$
$\hat{\mu}$
$\hat{m}$
2
1.5
1
0.5
2.5
2
1.5
1
0.5
0.2
0.15
0.1
1.5
1
0.5
0
-0.5
0
50
100
Number of trials

A
B
C
D
E
F
G
H
$s_{\hat{\sigma}}$
$s_{\hat{\mu}}$
1
0.5
0
0
50
100
Number of trials


A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
$\hat{\sigma}$
$\hat{\sigma}$
$\hat{k}$
$\hat{\mu}$
$\hat{\mu}$
1.5
1
0.5
0
-0.5
2.5
2
1.5
1
0.5
1
0
-1
0
50
100
Number of trials


A
B
C
D
Probability Distribution
Probability Distribution
0.4
0.2
0
0.4
0.2
0
0
1
2
0
1
2
3
-0.5
0
0.5
1
1.5
0
1
2
0
1
2
3
-0.5
0
0.5
1
1.5
$\hat{\sigma}$
$\hat{k}$
$\hat{\mu}$
$\hat{\sigma}$
$\hat{k}$
$\hat{\mu}$