A mixed filter algorithm for cognitive state estimation from simultaneously recorded continuous and binary measures of performance

Article *in* Biological Cybernetics · August 2008

DOI: 10.1007/s00422-008-0227-z · Source: PubMed

Published in final edited form as:
Biol Cybern. 2008 July ; 99(1): 1–14. doi:10.1007/s00422-008-0227-z.

A mixed filter algorithm for cognitive state estimation from simultaneously recorded continuous and binary measures of performance

M. J. Prerau,
Program in Neuroscience at Boston University, Boston, MA 02215, USA, e-mail: prerau@bu.edu, URL: http://people.bu.edu/prerau/

A. C. Smith,
Department of Anesthesiology and Pain Medicine, University of California at Davis, Davis, CA 95616, USA e-mail: annesmith@ucdavis.edu

U. T. Eden,
Program in Neuroscience at Boston University, Boston, MA 02215, USA, e-mail: prerau@bu.edu, URL: http://people.bu.edu/prerau/

M. Yanike,
Center for Neural Science, New York University, New York, NY 10003, USA, e-mail: yanike@cns.nyu.edu

W. A. Suzuki, and
Center for Neural Science, New York University, New York, NY 10003, USA, e-mail: yanike@cns.nyu.edu

E. N. Brown
Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital, Boston, MA 02114, USA

Department of Brain and Cognitive Sciences and the Massachusetts Institute of Technology/ Harvard Division of Health, Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, e-mail: brown@neurostat.mgh.harvard.edu

Abstract

Continuous (reaction times) and binary (correct/incorrect responses) measures of performance are routinely recorded to track the dynamics of a subject's cognitive state during a learning experiment. Current analyses of experimental data from learning studies do not consider the two performance measures together and do not use the concept of the cognitive state formally to design statistical methods. We develop a mixed filter algorithm to estimate the cognitive state modeled as a linear stochastic dynamical system from simultaneously recorded continuous and binary measures of performance. The mixed filter algorithm has the Kalman filter and the more recently developed recursive filtering algorithm for binary processes as special cases. In the analysis of a simulated learning experiment the mixed filter algorithm provided a more accurate and precise estimate of the cognitive state process than either the Kalman or binary filter alone. In the analysis of an actual learning experiment in which a monkey's performance was tracked by its series of reaction times, and correct and incorrect responses, the mixed filter gave a more complete description of the learning

Correspondence to: M. J. Prerau.

process than either the Kalman or binary filter. These results establish the feasibility of estimating cognitive state from simultaneously recorded continuous and binary performance measures and suggest a way to make practical use of concepts from learning theory in the design of statistical methods for the analysis of data from learning experiments.

Keywords

Cognitive state; Binary filter; Kalman filter; Learning; Gaussian approximation

1 Introduction

Learning is a dynamic process that is often defined as a change in behavior that occurs as consequence of experience. Learning is typically studied by showing that after a set of trial-and-error attempts, a subject is able to execute an unfamiliar task more reliably than chance would predict. The ability of an animal to learn a new task is frequently studied to characterize how attentional modulation (Cook and Maunsell 2002), genetics (Rondi-Reig et al. 2001), brain lesions (Roman et al. 1993; Whishaw and Tomie 1991; Dusek and Eichenbaum 1997; Fox et al. 2003; Wise and Murray 1999), or drug interventions (Stefani et al. 2003) affect learning. Accurate characterizations of learning are also important to relate changes in behavior to changes in neural activity in specific brain regions (Jog et al. 1999; Law et al. 2005; Wirth et al. 2003).

In a learning experiment, a subject is given a series of trials to learn an unfamiliar task. On each trial, the subject has a fixed amount of time to execute the task and his/her response is recorded as a 1 if the task is executed correctly and a 0 if it is executed incorrectly (Wirth et al. 2003). On each trial, continuous performance measures such as the subject's the reaction time or response time are also recorded. The reaction time is the amount of time that elapses from a go- or initiation-signal in the trial until the subject's response is first detected. The response time is the amount of time the subject takes to complete the task measured from the go-signal. Both the continuous and binary performance measures provide observations on the dynamics of the subject's cognitive state or understanding of the task that are believed to evolve with learning. Although both binary and continuous measures of performance are recorded in learning experiments the two are not used simultaneously to characterize cognitive state.

Simultaneous analysis of continuous and binary performance measures can give a more complete characterization of the cognitive state than either used separately. For example, in an experiment in which a subject has reached a plateau in his/her binary responses (i.e. gives all correct responses, yet continues to perform the task with increasingly faster response reaction times), examining solely the binary responses would suggest that learning has ceased, whereas the faster response times would suggest that the subject may be refining his/her understanding of the task. Therefore, the binary responses provide information on task acquisition, whereas the reaction times provide information on how the subject refines his/her understanding of the task.

To use the two types of observations simultaneously requires an explicit model of how these performance measures relate to the subject's cognitive state. Much research has been reported on models of cognitive state. These models have been used primarily to conduct theoretical characterizations of learning (Kakade and Dayan 2002; Luce 1965; Suppes 1959, 1990) and not to develop a statistical model to analyze the dynamics of cognitive states in learning experiments. The fact that a key objective of a learning experiment is to characterize an unobservable process, the cognitive state, from observable performance measures suggests that state-space modeling would be an appropriate analytic paradigm for studying this problem.

State-space modeling is an established framework widely used to study dynamical processes in engineering, computer science and statistics (Fahrmeir and Tutz 2002; Kitagawa and Gersch 1996; Mendel 1995; Roweis and Ghahramani 1999). Applications of this paradigm range from control systems (Kitagawa and Gersch 1996; Mendel 1995), speech recognition (Becchetti and Ricotti 1999), studies of protein structure (White et al. 1994), analysis of genetic networks (Harbison et al. 2004), studies of neural representations of biological signals (Brown et al. 1998; Barbieri et al. 2004; Eden et al. 2004; Smith and Brown 2003) and analyses of learning (Wirth et al. 2003; Smith et al. 2004). A state-space model consists of two equations: a state equation and an observation equation. The state equation defines an unobservable process whose evolution is tracked across time. Because the state process is unobservable, these models are often referred to as hidden Markov or latent process models (Fahrmeir and Tutz 2002; Roweis and Ghahramani 1999; Smith and Brown 2003; Smith et al. 2004). The observation equation defines how the state is measured or recorded. The objective of a state-space analysis is to estimate the states or the unobservable, hidden process from the sequence of observations. When the state and observation models are both linear and Gaussian, the well-known Kalman filter gives the optimal estimate of the unobserved states (Mendel 1995). Extensions of this filter to nonlinear system and/or nonlinear observation models have been widely studied using a range of computational strategies (Fahrmeir and Tutz 2002; Kitagawa and Gersch 1996; White et al. 1994; Doucet et al. 2001).

Combined analysis of continuous and binary observations has a long history in statistics (Hadidi and Schwartz 1979; Murphy 1999; Roumeliotis and Bekey 2000; Vlachonikolis 1990; Zeger and Liang 1986). More recently, state estimation from simultaneously continuous and binary observations has been studied using longitudinal data methods (Dunson 2003). Rudimentary state-space models have been used to analyze learning from binary observations alone (Smith et al. 2004, 2005, 2007). To date, state-space methods which combine information from continuous and binary performance measures and model explicitly the cognitive state have not been considered.

To devise such methods, we present a recursive filter algorithm to estimate a state-space model from simultaneously recorded continuous and binary observations. We term this algorithm a mixed filter. We illustrate the algorithm in the analysis of a simulated learning experiment in which the cognitive state of a subject is simultaneously observed through continuous and binary measures of performance, and in the analysis of an actual learning experiment in which a monkey performs a location-scene association task and its performance is tracked by both its sequence of reaction times and its sequence of correct and incorrect responses.

2 Theory and methods

In this section we formulate our mixed filter algorithm. Although the theory for our algorithm can be developed generally, we illustrate its essential features using a learning experiment with a one-dimensional state equation and an observation model for the continuous measurements and one for the binary measurements because this is the type of data we analyze in the Results. We assume that we are conducting a learning experiment and at each trial k, $k = 1, \dots, K$, the subject's ability to perform the task is governed by an unobservable cognitive state process x_k defined by the first-order autoregressive model with drift as

$$x_k=\rho_0+\rho x_{k-1}+\upsilon_k, \tag{1}$$

where the υ_k are independent, zero mean Gaussian random variables with variance σ_υ^2. Equation (1) is an elementary version of a cognitive state model developed by (Usher and McClelland 2001). As mentioned above, the cognitive state process specifies how the subject's learning or

understanding of the task at trial k is related to the subject's understanding of the task at trial $k-1$. The strength of that relation is given by the serial correlation coefficient ρ. The drift term ρ_0 defines a non-zero learning rate or propensity for the subject to learn the task. If $\rho_0 > 0$ then on average with repeated exposures to the task, the subject's cognitive state increases consistent with learning whereas if $\rho_0 < 0$, then on average, the subject's cognitive state will decline suggesting an inability to learn the task.

Let z_k and n_k denote respectively the continuous and the binary observations at time k. For each k, we assume $z_k \in (-\infty, \infty)$ and n_k is either 0 or 1. The observation model for the continuous performance measure is

$$z_k = \alpha + \beta x_k + \varepsilon_k, \tag{2}$$

where the ε_k are independent zero mean Gaussian random variables with variance σ_ε^2. We assume that the ε_k and the υ_k are independent. The parameter α governs the baseline reaction time, whereas β represents the rate at which the subject reacts as a function of his/her cognitive state. For an experiment in which a subject learns we would expect $\beta < 0$. We take $z_k = \log(r_k)$, where r_k is the reaction time at trial k. We assume this relationship because empirically reaction times typically decline rapidly as a subject first learns and later tend to reach a lower bound based on subject-specific physiological constraints. For example, there is a physical limit to how fast a subject can move his/her hand in response to a cue. The use of the logarithmic transformation with the Gaussian error assumption models the empirical observation that larger reaction times tend to show greater variability than shorter reaction times in learning experiments. By (2), this feature is captured by the fact that the distribution of the reaction times is log-normal. Hence, the value of the error is proportional to the value of the reaction time. Use of this logarithmic transformation also ensures that the reaction time estimates are always positive.

For the binary process we assume the Bernoulli observation model

$$\Pr(n_k|x_k) = p_k^{n_k}(1 - p_k)^{1-n_k}, \tag{3}$$

where n_k is 1 if the response is correct and 0 if it is incorrect, and p_k, the probability that the binary process takes the value 1 is defined in terms of the cognitive state process by the logistic equation

$$p_k = \exp\{\mu + \gamma x_k[\,1 + \exp\left\{\mu + \gamma x_k\right\}]^{-1}, \tag{4}$$

where γ is a modulation parameter which governs the effect of the cognitive state process on the probability of observing the binary outcome and μ defines the probability of the binary outcome when the state process is zero. We let $Z_k = [z_1, \ldots, z_k]$ and $N_k = [n_1, \ldots, n_k]$ be respectively the sequences of continuous and binary performance measures from trials 1 through k.

Our objective is to construct a recursive filter to estimate the state x_k at trial k from Z_k and N_k. The standard approach to deriving such a filter is to express recursively the probability density of the state given the observations. For the state and observation processes defined in (1)–(4), the probability density of the state given the continuous and binary observations is (Kitagawa and Gersch 1996)

$$p(x_k|Z_k,N_k)=\frac{p(x_k|Z_{k-1},N_{k-1})p(z_k|x_k)p(n_k|x_k)}{p(n_k,z_k|Z_{k-1},N_{k-1})}, \quad (5)$$

and the associated one-step prediction probability density or Chapman-Kolmogorov equation is

$$p(x_k|Z_{k-1},N_{k-1})= \int p(x_{k-1}|Z_{k-1},N_{k-1})p(x_k|x_{k-1})dx_{k-1}. \quad (6)$$

Before presenting the explicit form of the mixed filter algorithm, we explain the logic behind (5) and (6). The probability densities $p(z_k|x_k)$ and $p(n_k|x_k)$ in the numerator of (5) are respectively the Gaussian observation model for the reaction times defined in (2) and the Bernoulli observation model for the binary performance measures in defined (3) and (4). As is standard in state-space modeling, the numerator of (5) combines information from the one-step prediction of the state at trial k based on the observation up to through trial $k - 1$ given by the first term in the numerator and the two observation processes. The denominator is simply the normalizing constant of the probability density.

The one-step prediction density, $p(x_k|Z_{k-1}, N_{k-1})$ in (6) (Fig. 1) is the probability density of the state at trial k given the observations up through trial $k - 1$ and prior to recording continuous and binary responses at trial k. Equation (6) computes this probability density of the state at trial k by "averaging over" the state at trial $k - 1$ given the data up through trial $k - 1$ defined by $p(x_{k-1}|Z_{k-1}, N_{k-1})$ and the changes in state between trials $k - 1$ to k defined by $p(x_k|x_{k-1})$. The density of the states at trial $k - 1$ given the data up through $k - 1$, is the posterior density at $k - 1$, $p(x_{k-1}|Z_{k-1}, N_{k-1})$ which is the first term of the integrand in (6). The density for the change in state from trial $k - 1$ to k, is the second term in the integrand in (6) $p(x_k |x_{k-1})$, and is the autoregressive probability model for the cognitive state defined in (1).

Together, (5) and (6) define a recursion that can be used to compute the probability of the state given the observations (Fig. 1). Equations (5) and (6) are recursive because (6) uses the posterior probability density at trial $k - 1$, $p(x_{k-1}|Z_{k-1}, N_{k-1})$, to generate the one-step prediction probability density at $k - 1$, $p(x_k|Z_{k-1}, N_{k-1})$ which in turn, makes it possible to compute the new posterior probability at trial k, given in (5). As stated above, at each trial k, the recursion relations fuse information from two sources. The first is the prediction of the state at trial k given the model and the observations up through trial $k - 1$, as given in (6) and the second is the stochastic nature of the continuous and binary performance measures recorded at trial k defined respectively by the probability models (2)–(4).

In principle, given the model in (1)–(4) and the recursion relations in (5) and (6), the estimation of the cognitive state process from the performance measures is simply a computational problem. For systems with low-dimensional state and observation models, (5) and (6) can be evaluated numerically (Kitagawa and Gersch 1996). As the dimension of the system increases, numerical computation becomes less feasible. A standard approach, and the one we derive in the appendix and apply here, is to compute Gaussian approximations to (5) and (6) (Brown et al. 1998;Barbieri et al. 2004;Eden et al. 2004;Smith and Brown 2003). This approach which is also termed maximum a posteriori estimation (Mendel 1995), amounts to finding the maximum and curvature of (5) as a function of the state x_k. The maximum corresponds to the mode of the posterior density (A.7) and the curvature defines its variance (A.8). These two quantities are sufficient to compute the Gaussian approximation to (5) because a Gaussian probability density is defined completely by its mean (mode) and variance. Assuming a

Gaussian approximation to (5) was computed at trial $k - 1$, then the integral in (6) can be computed analytically using standard formulae for computing the mean and variance of the sum of two Gaussian random variables (A.9,A.10). Under the Gaussian approximation, the recursion defined by the probability densities in (5) and (6) becomes a recursive filter because it simplifies to computing recursively just the means and variances of these probability densities. In the special case that the state process is a linear Gaussian system and the observation model is a linear Gaussian function of the state process, this recursive computation of the means and variances is the Kalman filter (Fahrmeir and Tutz 2002;Kitagawa and Gersch 1996;Mendel 1995;Roweis and Ghahramani 1999). This would be true if our analysis did not include the binary performance measures.

Using the Gaussian approximations of (5) and (6) derived in the appendix, we obtain the following recursive mixed filter algorithm:

One-Step Prediction

$$x_{k|k-1} = \rho_0 + \rho x_{k-1|k-1} \tag{7}$$

One-Step Prediction Variance

$$\sigma^2_{k|k-1} = \rho^2 \sigma^2_{k-1|k-1} + \sigma^2_{\upsilon} \tag{8}$$

Gain Coefficient

$$C_k = (\beta^2 \sigma^2_{k|k-1} + \sigma^2_{\varepsilon})^{-1} \sigma^2_{k|k-1} \tag{9}$$

Posterior Mode

$$\begin{aligned} x_{k|k} = \quad & x_{k|k-1} + C_k[\beta(z_k - \alpha - \beta x_{k|k-1}) \\ & + \gamma \sigma^2_{\varepsilon}(n_k - p_{k|k})] \end{aligned} \tag{10}$$

Posterior Variance

$$\sigma^2_{k|k} = [(\sigma^2_{k|k-1})^{-1} + \gamma^2 p_{k|k}(1 - p_{k|k}) + (\sigma^2_{\varepsilon})^{-1} \beta^2],^{-1} \tag{11}$$

where we assume x_0 is given and the notation $x_{k|j}$ denotes the cognitive state at trial k given the observations up to and including trial j for $k = 1, \ldots, K$.

The algorithm in (7) to (11) is an analog of the Kalman filter for simultaneously recorded continuous and binary observations. For this reason, we term it a mixed filter algorithm. Equations (7) and (8) approximate (6) as the Gaussian probability density with mean $x_{k|k-1}$ and variance $\sigma^2_{k|k-1}$, whereas (10) and (11) approximate (5) as the Gaussian probability density with mean $x_{k|k}$ and variance $\sigma^2_{k|k}$. The relation between the cognitive state process, the observation models, the one-step-prediction density, the posterior probability density and their Gaussian approximations as defined by the mixed filter algorithm is shown in Fig. 1.

Because there are two observation processes, (10) has a continuous innovation term, $(z_k - \alpha - \beta x_{k|k-1})$ and a binary innovation, term $n_k - p_{k|k}$. As is true in the Kalman filter, the continuous

innovation compares the observation z_k with its one-step prediction. In contrast, the binary innovation compares the binary observation n_k with $p_{k|k}$, the probability of a correct response at trial k, where $p_{k|k} = [1 + \exp\{\mu + \gamma x_{k|k}\}]^{-1} \exp\{\mu + \gamma x_{k|k}\}$ by (4). As in the Kalman filter, C_k in (9), is a trial-dependent gain coefficient. At trial k, the amount by which the continuous innovation term affects the update is determined by $C_k\beta$, whereas the amount by which the binary innovation affects the update is determined by $C_k\gamma\sigma_\varepsilon^2$. The relative sizes of β and σ_ε^2 determine the relative weight given to the two innovation processes. For example, if σ_ε^2, the observation variance for the continuous performance measure is large relative to $\beta\gamma^{-1}$, then the binary innovation receives relatively more weight. Unlike in the Kalman filter algorithm, the left and right hand sides of the posterior mode (10) and the posterior variance (11) depend on the state estimate $x_{k|k}$. That is, because $p_{k|k} = [1 + \exp\{\mu + \gamma x_{k|k}\}]^{-1} \exp\{\mu + \gamma x_{k|k}\}$ is a function of $x_{k|k}$, the left and right sides of (10) are both functions $x_{k|k}$. Therefore, at each step k of the algorithm, we use Newton's methods to compute $x_{k|k}$ in (10). Either the previous state estimate, $x_{k-1|k-1}$, or the one-step prediction estimate, $x_{k|k-1}$, can provide a reliable starting guess. In our analyses, this Newton's procedure converges in two to four iterations.

If $\gamma = 0$, i.e., there is no relation between the binary performance measure and the cognitive state, then (7)–(11) simplify to a Kalman filter. On the other hand, if $\beta = 0$, so that there is no relation between the continuous performance measure and the cognitive state, then (7)–(11) simplify to the recursive filter algorithm for estimating an unobservable cognitive state process from binary performance measures given in (Smith and Brown 2003;Smith et al. 2004,2005).

3 Results

3.1 Mixed filter analysis of a simulated learning experiment

To test our mixed filter algorithm we simulated a learning experiment in which the cognitive state was modeled as a linear stochastic system observed simultaneously through a continuous performance measure and a binary performance measure. We modeled the cognitive state as the first-order autoregressive process with positive drift defined in (1). We chose the parameters for (12) to be $\rho = 0.99$, $x_0 = -3$, $\sigma_\nu^2 = 0.0886$, and $\rho_0 = 0.03$. For the continuous performance measure or reaction time process (2), we chose the parameters $\alpha = 1.0489$, $\beta = -0.2165$, and $\sigma_\varepsilon^2 = 0.4944$, and for the binary observation models or series of correct or incorrect responses, (4) we chose $\gamma = 0.99$ and $\mu = -0.5$. Because a state-space model of the form we proposed here has not been previously reported, we determined these parameter values for the simulated model based on preliminary analyses of performance data from the experiments described in (Wirth et al. 2003;Smith et al. 2004,2007). We learned from these analyses that parameter values such as the ones chosen here would generate performance data that were consistent with those reported when learning occurred.

Using (1), we first simulated $K = 200$ observations from the cognitive state process (Fig. 2b, c black curves). For each value of x_k, we used (2) to simulate the reaction times (Fig. 2a). For each value of the cognitive state process x_k we computed p_k from (4), and we simulated the sequence of correct and incorrect responses (Fig. 2d), by drawing a Bernoulli observation n_k with p_k defining the probability of a correct response.

We applied the mixed filter using the parameter values we chose to simulate the data, with the exception of ρ_0, which we took to be 0. We set $\rho_0 = 0$ in the mixed filter algorithm in (7) to make the analysis of the simulated data more realistic because the cognitive state model used to simulate the data is different from the one used to construct the filter for cognitive state estimation. By not including $\rho_0 > 0$ in the mixed filter algorithm, the analysis does not bias the algorithm toward an assumption of learning by forcing the cognitive state to increase on average with each trial. If $\rho_0 = 0$, then the cognitive state will only increase if both the observed binary

and continuous performance measures are consistent with learning. Furthermore, the choice of $\rho_0 = 0$ allows the mixed filter algorithm to identify other important behaviors seen in learning experiments such as failure to learn (Smith et al. 2007) as well extended periods of indecisive performance before performance consistent with learning appears (Smith et al. 2004).

We used Newton's method to compute (10) and (11) at each trial k with the starting value taken as $x_{k|k-1}$. We used a Kalman filter algorithm to estimate the cognitive state process and reaction time curve using just the reaction time measurements and we used the binary filter algorithm to estimate the cognitive state process and learning curve using just the binary responses. We then compared the Kalman and binary filter estimates of the cognitive state with the mixed filter state estimates computed by using the reaction time measurements and binary responses together. For the state estimate from each of the three filters, we computed the reaction time curve from (2) as $\exp\{\alpha + \beta x_{k|k}\}$ and the learning curve from (4) as $p_k = \exp\{\mu + \gamma x_{k|k}\} [1 + \exp\{\mu + \gamma x_{k|k}\}]^{-1}$.

We first analyzed the reaction times using a Kalman filter to estimate the cognitive states and their 95% confidence intervals (Fig. 2b) as well as the reaction time curve and its 95% confidence intervals (Fig. 2a) computed from the probability density in (A.12). The simulated observations were highly variable relative to the estimated reaction time curve and its confidence intervals (Fig. 2a). With the exception of a few trial sequences, such as at trials 1 to 3 and 14–18, the 95% confidence intervals covered the true cognitive state process (Fig. 2b, black curve). We next analyzed the binary observations using the binary filter algorithm (Wirth et al. 2003;Smith and Brown 2003;Smith et al. 2004,2005) to estimate the cognitive state, and its 95% confidence intervals (Fig. 2c), as well as the learning curve and its 95% confidence intervals (Fig. 2d). The confidence intervals were computed from the probability density in (A.13). The cognitive state estimated from the binary filter (Fig. 2c, gray curve) showed overshoots and undershoots relative to the true cognitive state process. In addition, the binary filter estimate of the cognitive state showed discrete declines when a sequence of correct responses was interrupted with an incorrect response (Fig. 2c), such as around trials 121, 139, and 167. The binary filter estimate of the learning curve (Fig. 2d) is highly consistent with the pattern of binary responses.

The cognitive state estimate from the mixed filter more closely resembled the true cognitive state process (Fig. 3b). The 95% confidence intervals from the mixed filter covered the true cognitive state process at nearly every trial (Fig. 3b). Although the mixed filter estimates of the reaction time (Fig. 3a) and learning curves (Fig. 3c) resembled their respective Kalman filter and binary filter estimates, the effect of the binary information on the Kalman estimates and vice versa were readily apparent. For example, the greater variations in the reaction time induced more fluctuations in the learning curve estimates and their 95% confidence intervals (Fig. 3c) at all trials compared to the binary filter estimates (Fig. 2d). At each trial, the mixed filter confidence intervals for the reaction time curve (Fig. 3a) are wider than the corresponding Kalman filter confidence intervals (Fig. 2a).

To analyze the performance characteristics of the mixed filter algorithm relative to the Kalman and binary filters, we simulated 1,000 realizations of continuous and binary measures of performance from a 200-trial learning experiment using the model in (1)–(4) with the parameters values described above. For each realization, we computed the mixed, Kalman, and binary filter estimates of the cognitive state. We compared performance of the algorithms in terms the mean-squared error, coverage probability for the 95% confidence intervals, and the median width of the 95% confidence intervals (Table 1). The mixed filter had the smallest mean-squared error (0.6001), the shortest median width of the 95% confidence intervals (3.0063), and the coverage probability for the 95% confidence intervals that was the closest to 0.95 (0.945).

These results show that the mixed filter algorithm can be both more accurate (more accurate coverage probability and narrower confidence intervals) and more precise (smaller mean squared error) in estimating the unobserved cognitive state, and that this filter can recover more information from the continuous and binary observations than state estimation with either of these sets of observations analyzed separately (Fig. 4).

3.2 Mixed filter analysis of an actual learning experiment

To illustrate the performance of our methods in the analysis of an actual learning experiment, we analyzed the responses of a macaque monkey in a location-scene association task, described in detail in (Wirth et al. 2003; Smith et al. 2004). In this task, the monkey fixated a point on a computer screen for a specified period and then, was presented with a novel scene. A delay period followed, and to receive a reward, the monkey had to associate the scene with the correct one of four target locations: north, south, east and west. Once the delay period ended, the monkey indicated its choice by making a saccadic eye movement to the chosen location. Because there were four locations the monkey could choose as a response, the probability of a correct response occurring by chance was 0.25 at the outset of the experiment. In addition to the sequence of correct and incorrect responses, the animal's performance on the 45-trial experiment was tracked trial-by-trial by its reaction times. In each trial a delay interval of 700 ms was imposed between the scene presentation and the response periods of the task. In this way, the animal was prevented from responding as soon as the scene presentation was given, obviating the possible trade off between the accuracy of the response and the speed of the task execution.

We used the EM algorithm (Dempster et al. 1997; Smith and Brown 2003; Smith et al. 2004, 2005; Shumway and Stoffer 1982), to estimate simultaneously all of the model parameters, including the initial condition x_0. We computed the reaction time and the learning curves, and their associated 95% confidence intervals from (A.12) and (A.13), respectively, in the appendix. The animal's reaction times were around 4.5 s for the first 21 trials and then declined linearly to approximately 2.5s by trial 45 (Fig. 5a). Consistent with this behavior, the monkey had only 3 correct responses in the first 25 trials and then it had all correct responses from trials 26 to 45 (Fig. 5d).

As in the analysis of the simulated learning experiment, we used a Kalman filter algorithm to estimate the cognitive state process and reaction time curve using just the reaction time measurements and we used the binary filter algorithm to estimate the cognitive state process and learning curve using just the binary responses. We then compared the Kalman and binary filter estimates of the cognitive state with the mixed filter cognitive state estimates computed by using the reaction time measurements and binary responses together. The estimated state processes from the Kalman (Fig. 5b) and binary filter (Fig. 5c) algorithms resembled each other with some differences. Because the animal gave incorrect responses from trial 2 to 20 followed by all correct responses from trial 26 to 45, the estimated cognitive state process from the binary filter initially declined monotonically then increased monotonically (Fig. 5c). In contrast, because the reaction times varied around a mean of approximately 4.5 sec for the first 20 trials and then declined in an approximate linear manner from trial 20 to trial 45 (Fig. 5a), the Kalman filter estimate of the cognitive state was initially constant then increased linearly (Fig. 5b). The learning curve had the same dynamics as the binary filter estimate of the cognitive state (Fig. 5d). Because $\beta = -0.03$ the reaction time curve was approximately constant up through trial 20 (Fig. 5a) then declined linearly as the Kalman estimate of the cognitive state increased from trial 21 to 45 (Fig. 5b). The effect of a single observation on the cognitive state and performance curve estimates can be seen at trials 6 and 30 for the reaction time observations and trials 21 and 24 for the binary observations (Fig. 5d).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The mixed filter estimates of the cognitive state process (Fig. 6b), the reaction time curve (Fig. 6a) and the learning curve (Fig. 6c) shared features of the Kalman and binary estimates. Like the binary filter estimate of the cognitive state (Fig. 5c), the mixed filter estimate of the cognitive state (Fig. 6b) declined monotonically from trials 1 to 20 because, although the reaction times were approximately constant over these trials, the effect of the 19 consecutive incorrect responses dominated the state estimation (Fig. 6c). As a consequence, unlike the Kalman filter reaction time curve estimate (Fig. 5a), the mixed filter reaction time curve estimate increased slightly over these trials (Fig. 6a). Similarly, the decline in reaction times beginning at trial 20 caused the mixed filter estimate of the cognitive state to increase beginning at that trial. For this reason, the learning curve also began to increase at this trial even though the animal had 3 of the 5 incorrect responses from trials 20 to 24 (Fig. 6c). Because the monkey has all correct responses from trial 26 to 45 and the reaction times decrease linearly over these trials (Fig. 6a), the mixed filter estimates of the cognitive state (Fig. 6b) and learning curve (Fig. 6c) increased and the reaction time curve (Fig. 6a) decreased.

The outliers in the reaction times affected the learning curve estimates. The short reaction time at trial 6 and the long reaction time at trial 33 (Fig. 6a) caused a noticeable increase in the width of the learning curve confidence intervals on those trials (Fig. 6c). This is because these reaction times were inconsistent with the binary responses immediately preceding and following those trials. The declining reaction times between trials 20 and 24 (Fig. 6a) were responsible for the increase in the learning curve over these trials even though, the animal had mostly incorrect responses (Fig. 6c). This brief inconsistency between the reaction time and the binary responses was reflected in the increased width of the learning curve confidence intervals over these trials. Similarly, the short reaction time at trial 30 during an interval in which all of the animal's responses were correct, led to a sharp decrease in the width of the learning curve confidence intervals.

Unlike an analysis which presupposed improvement during a learning experiment and therefore, used a learning model that increased monotonically on average as a function of trial number in the experiment, i.e. assumes $\rho > 0$ (Granat et al. 2003), the dynamic nature of the cognitive state model with $\rho = 0$ allowed it to capture a range of behaviors during the learning experiment. In the current experiment, this flexibility was clearly illustrated by the estimated cognitive state that initially declined in the first half of the experiment, then increased non-monotonically in the latter half.

Plotting the Kalman, binary and mixed filter estimates of the cognitive state on the same graph further illustrates how the mixed filter estimate combines features of the other two (Fig. 7). The mixed filter estimate of the cognitive state resembled the fine structure of the Kalman filter estimate, whereas it was more closely associated with the level of the binary filter estimate of the cognitive state. These results show that the mixed filter analysis gave a more informative characterization of performance in this learning experiment compared with either a separate Kalman filter or binary filter analysis.

4 Discussion

We have developed a mixed filter algorithm using the state-space paradigm to track cognitive state dynamics in learning experiments from simultaneously recorded continuous and binary measures of performance. The mixed filter algorithm has the standard Kalman filter and the recursive algorithm for filtering binary processes as special cases. We illustrated the algorithm, which uses an elementary model of learning, in the analysis of both simulated and actual learning experiments. In the simulated example, the mixed filter algorithm provided a more accurate (greater coverage probability and narrower confidence intervals) and more precise (smaller means squared error) estimates of the cognitive state process than either the Kalman

filter or the binary filter. In the analysis of the actual data, the mixed filter cognitive state estimates clearly combined features of the other two estimates, and gave a more complete description of the learning process during the experiment. Formal use of the cognitive state concept in the analyses of learning experiments should help relate more directly findings in experimental and theoretical studies of learning.

In learning studies using only the binary performance measures the binary filter algorithm is used in conjunction with the corresponding state-space smoothing algorithm to conduct the data analysis (Wirth et al. 2003; Smith et al. 2004, 2005; Prerau et al. 2004). Constructing the corresponding smoothing algorithm for our mixed filter algorithm is a straight forward extension of the current work. Learning analyses that use continuous and binary performance measures will lead to new definitions of current quantitative measures of learning defined for only binary data analyses, such as the learning trial, bivariate comparison matrices, and the ideal observer curve (Smith et al. 2004, 2005, 2007). The extension of our new algorithm to a hierarchical framework would also make it possible to combine information across subjects in learning experiments using both continuous and binary measures of performance (Smith et al. 2007).

While this work suggests the feasibility of using a state-space model to analyze the cognitive state of a subject in a learning experiment from simultaneously recorded continuous and binary performance measures, several technical improvements are possible. First, the observation and cognitive state model can be extended to consider more complex models of the cognitive state and the relation between cognitive state and the performance measures. For example, the performance measures may be correlated rather than conditionally independent as assumed in the current analysis. As another possibility, the cognitive state may be more accurately represented as a multivariate process with nonmonotonic relations between the components of the cognitive state and the performance measures (Smith et al. 2007; Usher and McClelland 2001).An appropriate mixed filter algorithm could be derived for these more complex models using the approach we have described. If these new state-space models are nonlinear, then sequential Monte Carlo methods could be used to estimate the cognitive states (Doucet et al. 2001; Ergun et al. 2007).

Second, we used the standard Gaussian approximation to the posterior probability density to construct a simple, computationally tractable algorithm. Our experience has shown that in analyses of learning experiments that use just the binary filter algorithm, the posterior densities can be well approximated by a Gaussian (Anne Smith, personal communication). Further theoretical and empirical studies are needed to understand whether this Gaussian approximation accurately captures the true posterior density for a broad range of models of learning experiments in which continuous and binary performances are used simultaneously. These studies should consider factors such as the continuous and binary observation signal-to-noise ratios, the cognitive state signal-to-noise ratio, and the magnitude of observation modulations parameters. For cases in which the posterior is non-Gaussian but remains unimodal, this approximation may still accurately capture the behavior of the mode and variance. When the Gaussian approximation is no longer appropriate, it still may be useful for helping to compute the posterior more accurately by serving as a proposal density in a sequential Monte Carlo filter algorithm (Doucet et al. 2001; Ergun et al. 2007).

In addition to further use in the analysis of learning experiments, the mixed filter algorithm may be applicable to other biological and engineering problems in which combining information from continuous and binary time-series measurements is of interest. First, in neurophysiology experiments multiple neural spike trains are simultaneously recorded with local field potentials as a common approach to study how neural systems represent relevant biological signals. These data can be analyzed with decoding algorithms, i.e. state-space

NIH-PA Author Manuscript

NIH-PA Author Manuscript

estimation procedures, by extending the current algorithm to construct mixed filter algorithms for continuous observations and point processes (continuous time binary processes) in either discrete or continuous time (Eden et al. 2004; Snyder and Miller 1991). Second, the mixed filter algorithm may make it possible to use simultaneously recorded ensemble neural spiking activity and local field potentials to control neural prosthetic devices and brain machine interfaces (Musallam et al. 2004; Serruya et al. 2002; Taylor et al. 2002; Wessberg et al. 2005). Finally, the mixed filter algorithms may also suggest a new approach to analyzing cardiovascular and autonomic control from simultaneously recorded cardiovascular, respiratory and $R-R$ interval measurements (Barbieri et al. 1996, 1997, 2002) as well as for studying the dynamics of seismic events (Granat et al. 2003).

Acknowledgments

This work was supported in part by NIDA grant DA015644 to E.N.B and W.A.S, NIMH grant MH071847 to E.N.B. and NIMH grant MH58847 and a McKnight Foundation grant to W.A.S.

Appendix

Appendix A

A.1 Derivation of the mixed filter algorithm

To approximate a unimodal probability density $f(x)$ with a Gaussian probability density we compute its mode $\hat{\mu}$ as the solution to the equation (Tanner 1996)

$$\frac{\partial \log f(x)}{\partial x}\Big|_{\hat{\mu}}=0, \tag{A.1}$$

and its variance $\hat{\sigma}^2$ as

$$\hat{\sigma}^2=\left[-\frac{\partial^2 \log f(x)}{\partial x^2}\Big|_{\hat{\mu}}\right]^{-1}. \tag{A.2}$$

We derive the mixed filter algorithm by computing a Gaussian approximation to the posterior density $p(x_k|Z_k, N_k)$ in (5), (Brown et al. 1998;Barbieri et al. 2004;Eden et al. 2004;Smith and Brown 2003). At trial k, we assume that the one-step prediction probability density in (6) is the Gaussian probability

$$\begin{aligned} p(x_k|Z_{k-1},N_{k-1})= \quad & (2\pi\sigma^2_{k|k-1})^{-\frac{1}{2}}\exp\Big\{-(2\sigma^2_{k|k-1})^{-1} \\ & \times (x_k-\rho_0-\rho x_{k|k-1})^2\Big\}. \end{aligned} \tag{A.3}$$

The probability density for the continuous performance measures is

$$\begin{aligned} p(z_k|x_k)= \quad & (2\pi\sigma^2_{\varepsilon})^{-\frac{1}{2}}\exp\{-(2\sigma^2_{\varepsilon})^{-1} \\ & \times (z_k-\alpha-\beta x_k)^2\Big\}, \end{aligned} \tag{A.4}$$

and the probability mass function for the binary performance measures is given in (3). Substituting (A.3), (A.4) and (3) into (5) gives respectively the posterior probability density and the log posterior probability density as

$$p(x_k|Z_k,N_k) \propto \exp\Big\{-(2\sigma^2_{k|k-1})^{-1} \times (x_k - \rho_0 - \rho x_{k|k-1})^2 + n_k \log[p_k(1-p_k)^{-1}] + \log(1-p_k) - (2\sigma^2_\varepsilon)^{-1}(z_k - \alpha - \beta x_k)^2\Big\} \quad (A.5)$$

$$\log p(x_k|Z_k,N_k) = -(2\sigma^2_{k|k-1})^{-1}(x_k - \rho_0 - \rho x_{k|k-1})^2 + n_k \log p_k(1-p_k)^{-1} + \log(1-p_k) - (2\sigma^2_\varepsilon)^{-1}(z_k - \alpha - \beta x_k)^2. \quad (A.6)$$

To compute the maximum-a-posteriori estimate of x_k and its associated variance estimate, we compute the first and second derivatives of the log posterior probability density with respect to x_k, which are respectively

$$\frac{\partial \log p(x_k|Z_k,N_k)}{\partial x_k} = -(\sigma^2_{k|k-1})^{-1}(x_k - \rho_0 - \rho x_{k|k-1}) + (\sigma^2_\varepsilon)^{-1}\beta(z_k - \alpha - \beta x_k) + \gamma(n_k - p_k) \quad (A.7)$$

$$\frac{\partial^2 \log p(x_k|Z_k,N_k)}{\partial x_k^2} = -(\sigma^2_{k|k-1})^{-1} - (\sigma^2_\varepsilon)^{-1} \times \beta^2 - \gamma^2 p_k(1-p_k). \quad (A.8)$$

By (A.1) the mode (mean) of the Gaussian approximation to $p(x_k|Z_k, N_k)$ is obtained by setting (A.7) equal to zero and solving for x_k. This gives the posterior mode or maximum-a-posteriori estimate of x_k in (10). By (A.2), the variance of the Gaussian approximation to $p(x_k|Z_k, N_k)$ is the negative inverse of (A.8), and gives the posterior variance in (11). Equations (7) and (8) follow because given $x_{k-1|k-1}$ and using (1), we have that

$$x_{k|k-1} = E(x_k|x_{k-1|k-1}) = \rho_0 + \rho x_{k-1|k-1}, \quad (A.9)$$

and

$$\sigma^2_{k|k-1} = \mathrm{Var}(x_k|x_{k|k-1}) = \mathrm{Var}(\rho x_{k-1} + v_k|x_{k|k-1}) = \rho^2\sigma^2_{k-1|k-1} + \sigma^2_v. \quad (A.10)$$

A.2 Probability densities of the reaction time and time curves

In the learning experiment we modeled the continuous performance measure as $z_k = \log r_k = \alpha + \beta x_k + \varepsilon_k$. To compute the probability density of the reaction time r_k from the posterior probability density of the state process x_k we use the standard change-of-variables formula (Smith et al. 2004)

$$f(r)=f(x)\left|\frac{dx}{dr}\right|, \tag{A.11}$$

where $f(x_k)$ is the Gaussian probability density with mean $x_{k|k}$ and variance $\sigma^2_{k|k}$ and $\frac{dx}{dr}=(\beta r)^{-1}$. Applying (A.11) by substituting $x = \beta^{-1}$ (log $r - \alpha$) into $f(x)$ yields the lognormal probability density for the reaction times defined as

$$\begin{aligned} f(r|,x_{k|k},\sigma^2_{k|k})= \quad & [r\beta(2\pi\sigma^2_{k|k})^{\frac{1}{2}}]^{-1}\exp\{-(2\sigma^2_{k|k}\beta^2)^{-1} \\ & \times(\log(r-\alpha)-x_{k|k})^2\Big\}. \end{aligned} \tag{A.12}$$

Again, from the posterior probability density of any x_k, we can also compute the probability density of p_k using (3) and (4), and the change-of-variable formula in (A.11). In doing so, we set $x = \gamma^{-1}[\log[p(1 - p)^{-1}] - \mu]$ in $f(x)$ and replace $\frac{dx}{dr}$ with $\frac{dx}{dp} = [\gamma p(1 - p)]^{-1}$ to obtain

$$\begin{aligned} f(p|\mu,x_{k|k},\sigma^2_{k|k})= \quad & [\gamma p(1-p)(2\pi\sigma^2_{k|k})^{\frac{1}{2}}]^{-1} \\ & \times\exp\Big\{-(2\sigma^2_{k|k})^{-1}(\gamma[\log p[(1-p) \\ & \times\exp(\mu)]^{-1}]^{-1}-x_{k|k})^2\Big\}. \end{aligned} \tag{A.13}$$

We use (A.12) and (A.13) to compute 95% confidence intervals for the reaction time and learning curves respectively.

References

Barbieri R, Bianchi AM, Triedman JK, Mainardi LT, Cerutti S, Saul JP. Model dependency of multivariate autoregressive spectral analysis. IEEE Eng Med Biol Mag 1997 September-October; 16:74–85. [PubMed: 9313084]

Barbieri R, Frank LM, Nguyen DP, Quirk MC, Solo V, Wilson MA, Brown EN. Dynamic analyses of information encoding in neural ensembles. Neural Comput 2004 February;16:277–307. [PubMed: 15006097]

Barbieri R, Triedman JK, Saul JP. Heart rate control and mechanical cardiopulmonary coupling to assess central volume: a systems analysis. Am J Physiol Regul Integr Comp Physiol 2002 November; 283:R1210–R1220. [PubMed: 12376415]

Barbieri R, Waldmann A, DiVirgilio V, Triedman JK, Bianchi AM, Cerutti S, Saul JP. Continuous quantification of baroreflex and respiratory control of heart rate by use of bivariate autoregressive techniques. Ann. Noninvasive Electrocardiol 1996;1:264–277.

Becchetti, C.; Ricotti, LP. Speech recognition: theory and C++ implementation. Chichester: Wiley; 1999.

Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. J Neurosci 1998 September;18:7411–7425. [PubMed: 9736661]

Cook EP, Maunsell JH. Attentional modulation of behavioral performance and neuronal responses in middle temporal and ventral intraparietal areas of macaque monkey. J Neurosci 2002 March;22:1994–2004. [PubMed: 11880530]

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm (with discussion). J Roy Statist Soc B 1977;39:1–38.

Doucet, A.; De Freitas, N.; Gordon, N. Sequential Monte Carlo methods in practice. New York: Springer; 2001.

Dunson DB. Dynamic latent trait models for mutlidimensional longitudinal data. JASA 2003;98:555–563.

Dusek JA, Eichenbaum H. The hippocampus and memory for orderly stimulus relations. Proc Natl Acad Sci USA 1997 June;94:7109, 7114. [PubMed: 9192700]

Eden UT, Frank LM, Barbieri R, Solo V, Brown EN. Dynamic analysis of neural encoding by point process adaptive filtering. Neural Comput 2004 May;16:971–998. [PubMed: 15070506]

Ergun A, Barbieri R, Eden UT, Wilson MA, Brown EN. Construction of point process adaptive filter algorithms for neural systems using sequential Monte Carlo methods. IEEE Trans Biomed Eng 2007 March;54:419–428. [PubMed: 17355053]

Fahrmeir, L.; Tutz, G. Multivariate Statistical Modeling Based on Generalized Linear Models 2ed. New York, NY: Springer-Verlag; 2002.

Fox MT, Barense MD, Baxter MG. Perceptual attentional set-shifting is impaired in rats with neurotoxic lesions of posterior parietal cortex. J Neurosci 2003 January;23:676–681. [PubMed: 12533627]

Granat, RA.; Clayton, R.; Kedar, S.; Kaneko, Y. Regularized deterministic annealing hidding Markov models for identification and analysis of seismic and aseismic events. American Geophysical Union Fall Meeting; 2003.

Hadidi M, Schwartz S. Linear recursive state estimators under uncertain observations. IEEE Transactions on Automatic Control 1979;24:944–948.

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA. Transcriptional regulatory code of a eukaryotic genome. Nature 2004 September;431:99–104. [PubMed: 15343339]

Jog MS, Kubota Y, Connolly CI, Hillegaart V, Graybiel AM. Building neural representations of habits. Science 1999 November;286:1745–1749. [PubMed: 10576743]

Kakade S, Dayan P. Acquisition and extinction in autoshaping. Psychol Rev 2002 July;109:533–544. [PubMed: 12088244]

Kitagawa, G.; Gersch, W. Smoothness Priors Analysis of Time Series. New York, NY: Springer-Verlag; 1996.

Law JR, Flanery MA, Wirth S, Yanike M, Smith AC, Frank LM, Suzuki WA, Brown EN, Stark CE. Functional magnetic resonance imaging activity during the gradual acquisition and expression of paired-associate memory. J Neurosci 2005 June;25:5720–5729. [PubMed: 15958738]

Luce D. A theoretical analysis of detection in temporally unstructured experiments. Z Psychol Z Angew Psychol 1965;171:57–68. [PubMed: 4379746]

Mendel, JM. Lessons in Estimation Theory for Signal Processing, Communications, and Control. Englewood Cliffs, NJ: Prentice Hall; 1995.

Murphy KP. A variational approximation for Bayesian networks with discrete and continuous latent variables. UAI. 1999

Musallam S, Corneil BD, Greger B, Scherberger H, Andersen RA. Cognitive control signals for neural prosthetics. Science 2004 July;305:258–262. [PubMed: 15247483]

Prerau MJ, Kubota Y, Graybiel AM, Smith AC, Brown EN. Estimating learning from simultaneously recorded continuous and discrete measures. Society for Neuroscience Abstract. 2004

Roman FS, Simonetto I, Soumireu-Mourat B. Learning and memory of odor-reward association: selective impairment following horizontal diagonal band lesions. Behav Neurosci 1993 February;vol. 107:72–81. [PubMed: 8447959]

Rondi-Reig L, Libbey M, Eichenbaum H, Tonegawa S. CA1-specific N-methyl-D-aspartate receptor knockout mice are deficient in solving a nonspatial transverse patterning task. Proc Natl Acad Sci USA 2001 March;98:3543–3548. [PubMed: 11248114]

Roumeliotis, SI.; Bekey, GA. Bayesian estimation and Kalman filtering: a unified framework formobile robot localization. Proceedings of ICRA ’00. IEEE International conference on robotics and automation; 2000. p. 2985-2992.

Roweis S, Ghahramani Z. A unifying review of linear gaussian models. Neural Comput 1999 February; 11:305–345. [PubMed: 9950734]

Serruya MD, Hatsopoulos NG, Paninski L, Fellows MR, Donoghue JP. Instant neural control of a movement signal. Nature 2002 March;416:141–142. [PubMed: 11894084]

Shumway RH, Stoffer DS. An approach to time series smoothing and forecasting using the EM algorithm. Journal of Time Series Analysis 1982;3:253–264.

Smith AC, Brown EN. Estimating a state-space model from point process observations. Neural Comput 2003 May;15:965–991. [PubMed: 12803953]

Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki WA, Brown EN. Dynamic analysis of learning in behavioral experiments. J Neurosci 2004 January;24:447–461. [PubMed: 14724243]

Smith AC, Stefani MR, Moghaddam B, Brown EN. Analysis and design of behavioral experiments to characterize population learning. J Neurophysiol 2005 March;93:1776–1792. [PubMed: 15456798]

Smith AC, Wirth S, Suzuki WA, Brown EN. Bayesian analysis of interleaved learning and response bias in behavioral experiments. J Neurophysiol 2007 March;97:2516–2524. [PubMed: 17182907]

Stefani MR, Groth K, Moghaddam B. Glutamate receptors in the rat medial prefrontal cortex regulate set-shifting ability. Behav Neurosci 2003 August;117:728–737. [PubMed: 12931958]

Snyder, DL.; Miller, MI. Random point processes in time and space. Vol. 2nd edn. New York: Springer; 1991.

Suppes, P. A linear model for a continuum of responses. In: Bush, RR.; Estes, KK., editors. Studies in mathematical learning theory. Stanford, CA: Stanford University Press; 1959. p. 400-414.

Suppes P. On deriving models in the social-sciences. Math Comput Model 1990;14:21–28.

Tanner, MA. Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions. Vol. 3rd edn. New York: Springer; 1996.

Taylor DM, Tillery SI, Schwartz AB. Direct cortical control of 3D neuroprosthetic devices. Science 2002 June;296:1829–1832. [PubMed: 12052948]

Usher M, McClelland JL. The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev 2001 July;108:550–592. [PubMed: 11488378]

Vlachonikolis IG. Predictive discrimination and classification with mixed binary and continuous variables. Biometrika 1990;77:657–662.

Wessberg J, Stambaugh CR, Kralik JD, Beck PD, Laubach M, Chapin JK, Kim J, Biggs SJ, Srinivasan MA, Nicolelis MA. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. Nature 2005 November;408:361–365. [PubMed: 11099043]

Wise SP, Murray EA. Role of the hippocampal system in conditional motor learning: mapping antecedents to action. Hippocampus 1999;9:101–117. [PubMed: 10226772]

Whishaw IQ, Tomie JA. Acquisition and retention by hippocampal rats of simple, conditional, and configural tasks using tactile and olfactory cues: implications for hippocampal function. Behav Neurosci 1991 December;105:787–797. [PubMed: 1777102]

Wirth S, Yanike M, Frank LM, Smith AC, Brown EN, Suzuki WA. Single neurons in the monkey hippocampus and learning of new associations. Science 2003 June;300:1578–1581. [PubMed: 12791995]

White JV, Stultz CM, Smith TF. Protein classification by stochastic modeling and optimal filtering of amino-acid sequences. Math Biosci 1994 January;119:35–75. [PubMed: 8111135]

Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986 March;42:121–130. [PubMed: 3719049]

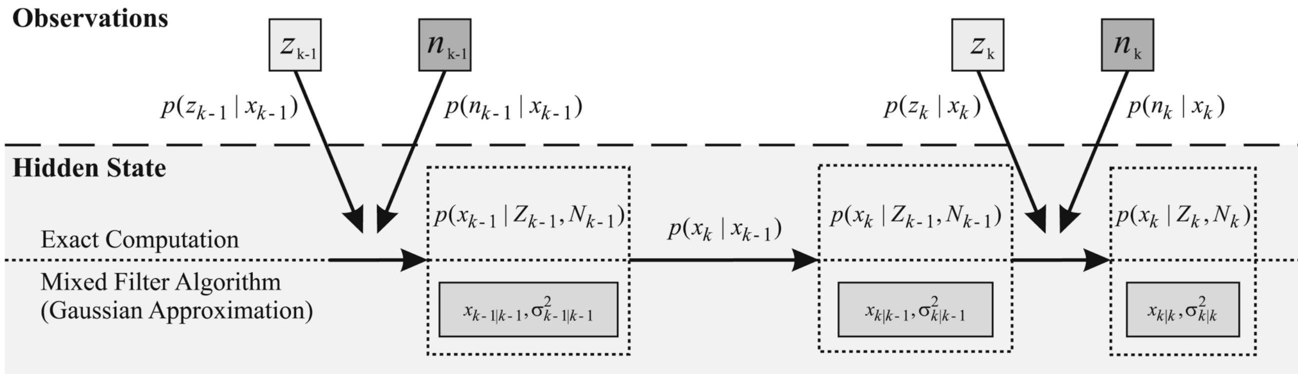

Fig. 1.

Schematic representation of the relation between the observation processes (continuous and binary performance measures) $p(z_k|x_k)$ (2), $p(n_k|x_k)$ (3), the cognitive state process $p(x_k|x_{k-1})$ (1), the one-step prediction density $p(x_k|Z_{k-1}, N_{k-1})$ (6), its mixed filter Gaussian approximation with mean $x_{k|k-1}$ (7) and variance $\sigma^2_{k|k-1}$ (8), and the posterior probability density $p(x_k|Z_k, N_k)$ (5), and its mixed filter Gaussian approximation with mean $x_{k|k}$ (10) and variance $\sigma^2_{k|k}$ (11)

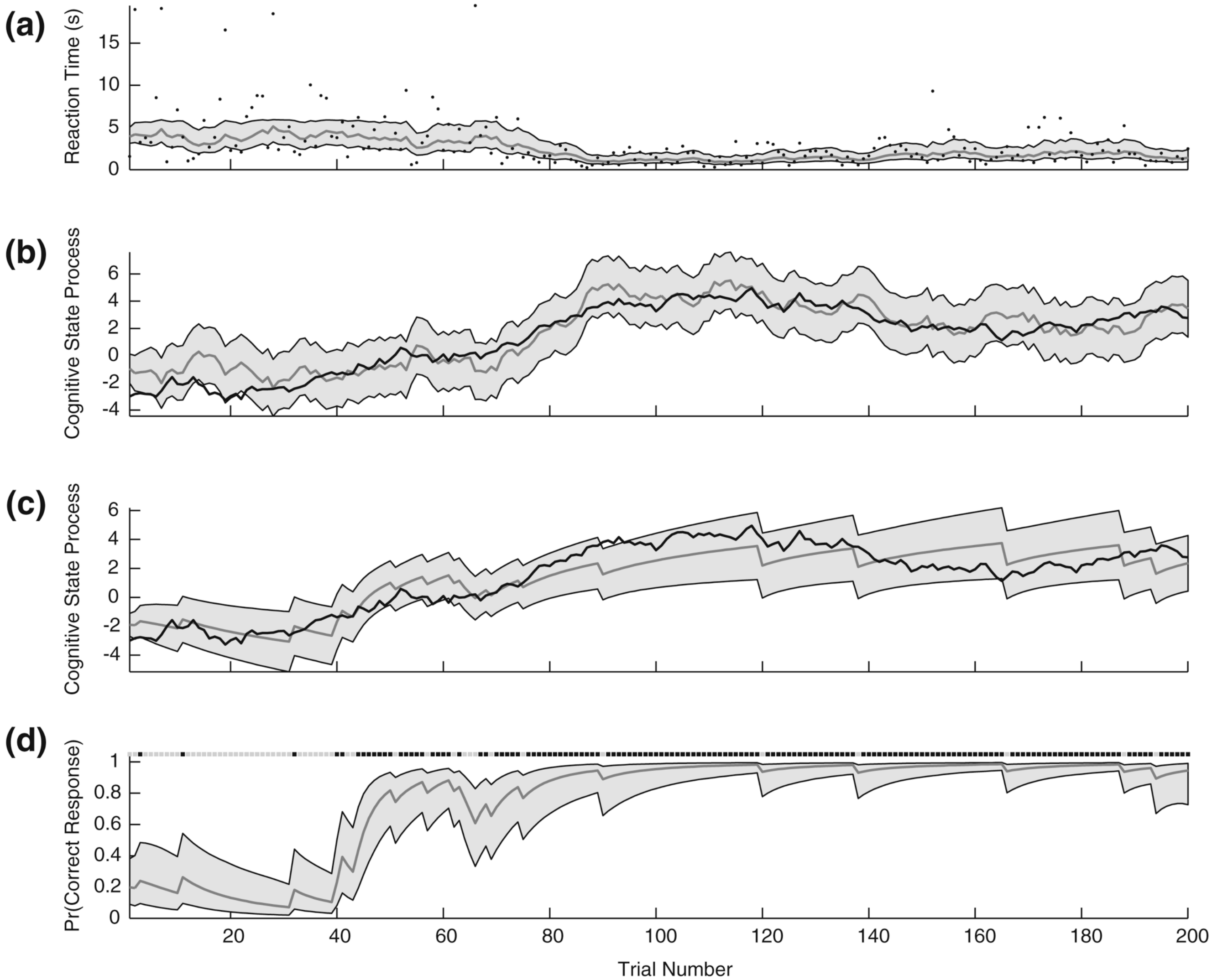

Fig. 2.

Kalman filter and binary filter analysis of the simulated learning experiments. Kalman filter estimate of the **a** reaction time curve (*gray*) and **b** the cognitive state process (*gray*), binary filter estimate of the **c** cognitive state process (*gray*) and **d** learning curve (*gray*) from the simulated learning experiment are in gray. *The light gray regions* are the 95% confidence intervals. The cognitive state process is shown as **a** black curve in **b** and **c**. *Black dots* in **a** are the reaction times. Correct (*black*) and incorrect (*gray*) trial responses are shown above **d**

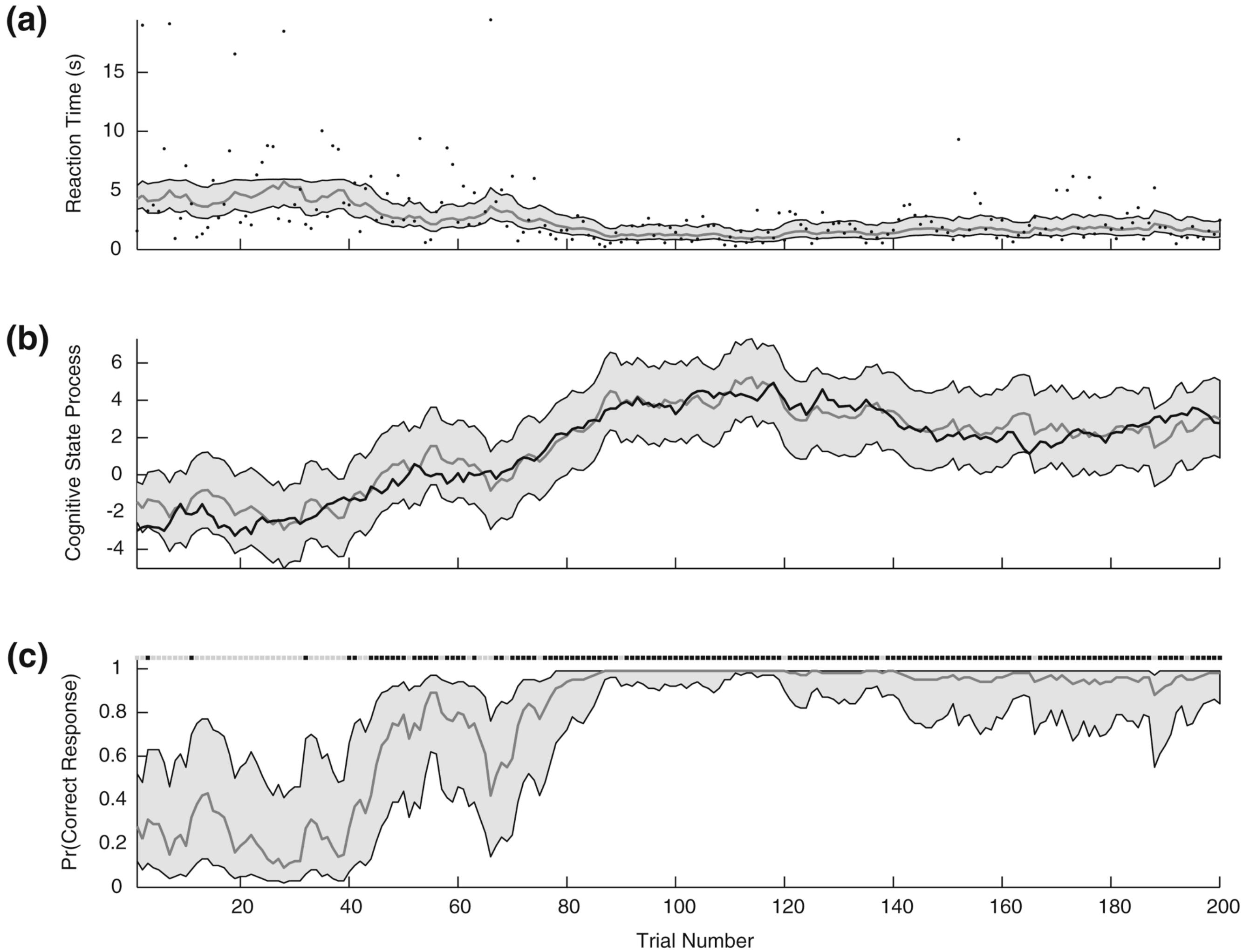

Fig. 3.

Mixed filter analysis of the simulated learning experiment. **a** Mixed filter estimate of the reaction time curve (*gray*). *Black* dots are the reaction times. **b** Mixed filter estimate of the cognitive state process (*gray*) and the true cognitive state process (*black*). **c** Mixed filter estimate of the learning curve (*gray*). Correct (*black*) and incorrect (*gray*) trial responses are shown above **c**. *The light gray* regions are the 95% confidence intervals in each panel

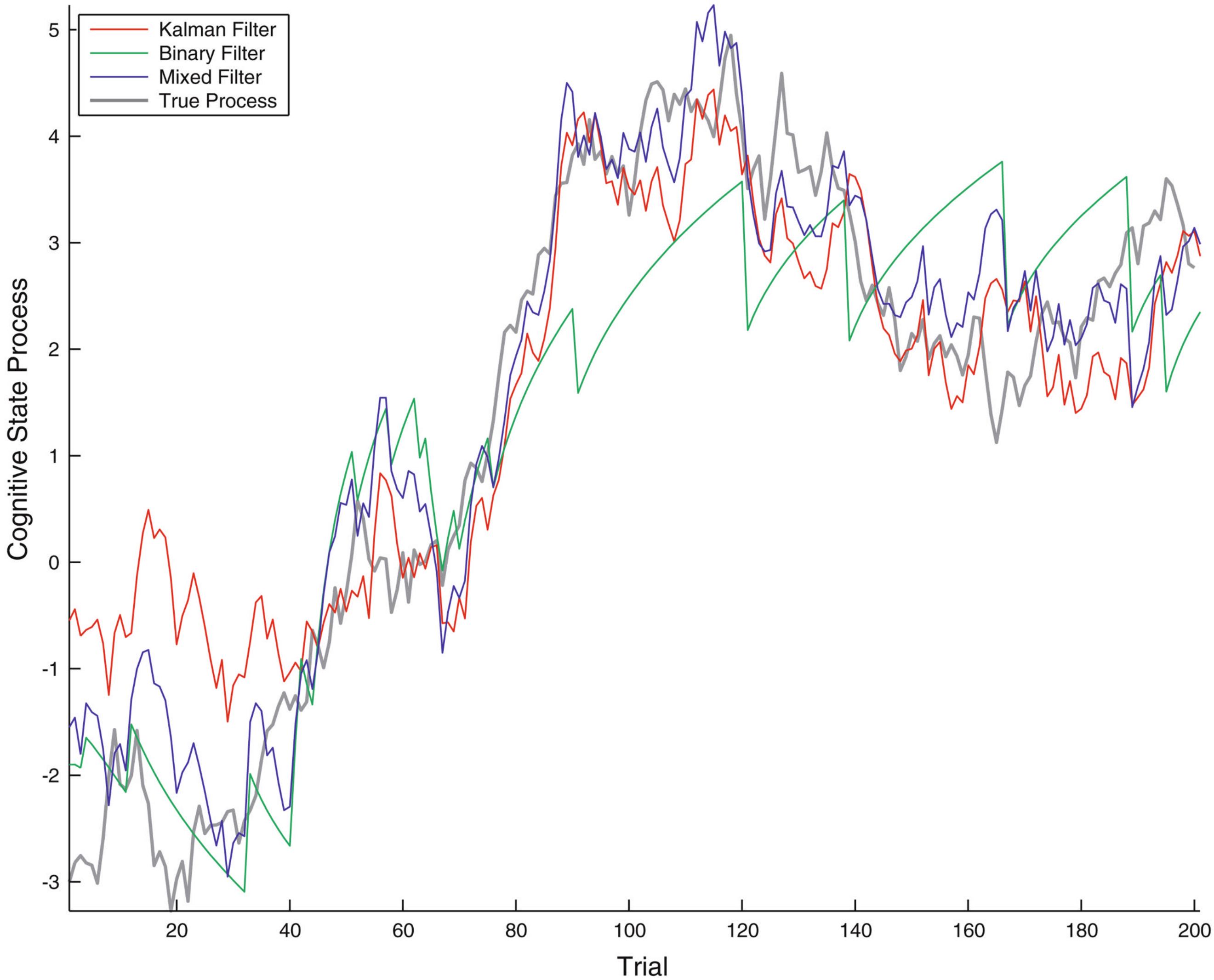

Fig. 4.
Binary filter (*green*), Kalman filter (*red*) and mixed filter (*blue*) estimates of the cognitive state process, along with the true cognitive state process (*thick gray*)

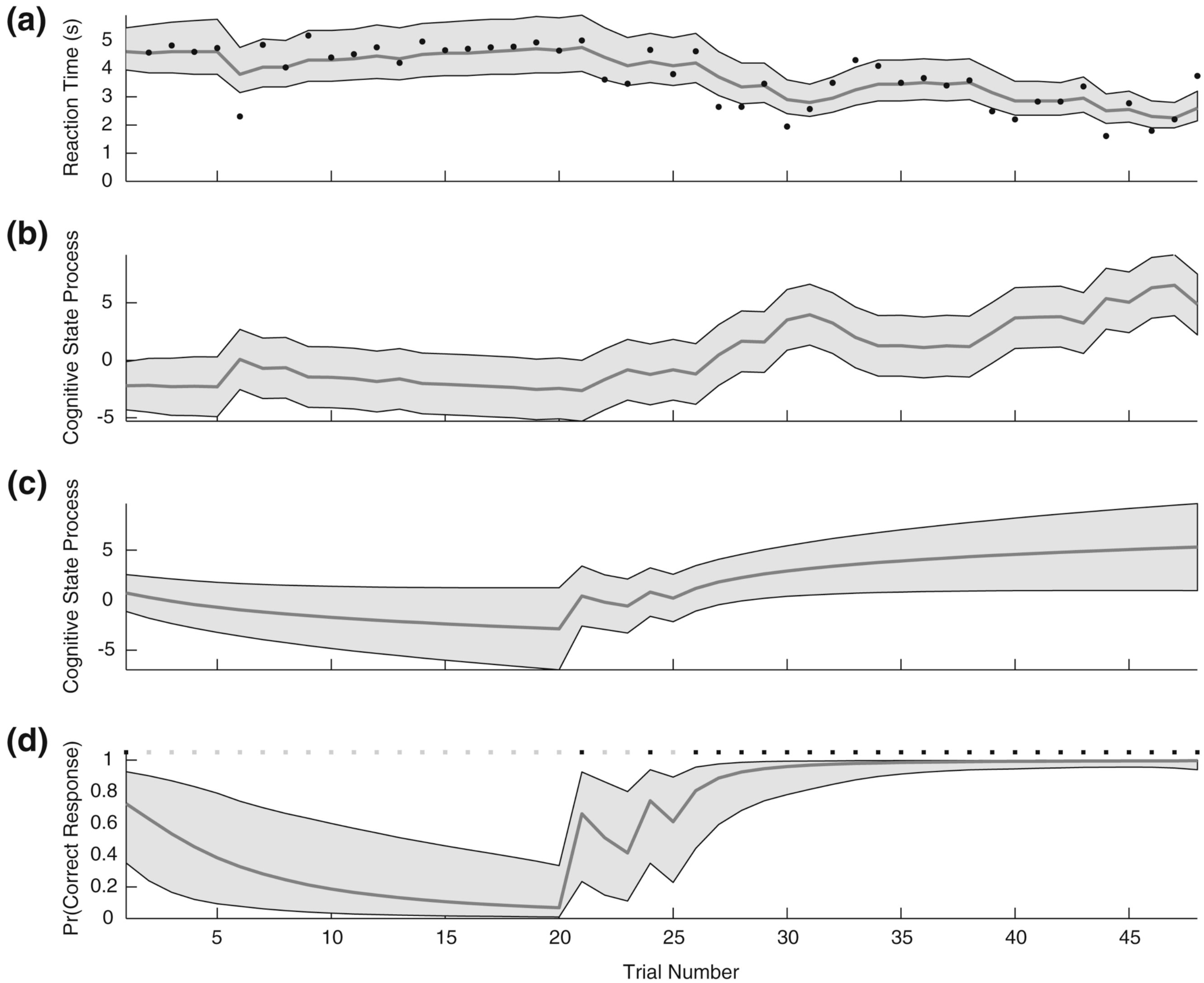

Fig. 5.

Kalman filter and binary filter analyses of a monkey's performance executing a location-scene association task. Kalman filter estimate of the reaction time curve (*gray*) **a** and the cognitive state process (*gray*). **b**. Binary filter estimates of the cognitive state process (*gray*) **c** and the learning curve (*gray*) **d**. The light gray regions are the 95% confidence intervals. *Black dots* in a are the measured reaction times. Correct (*black*) and incorrect (*gray*) trial responses of the monkey are shown above **d**

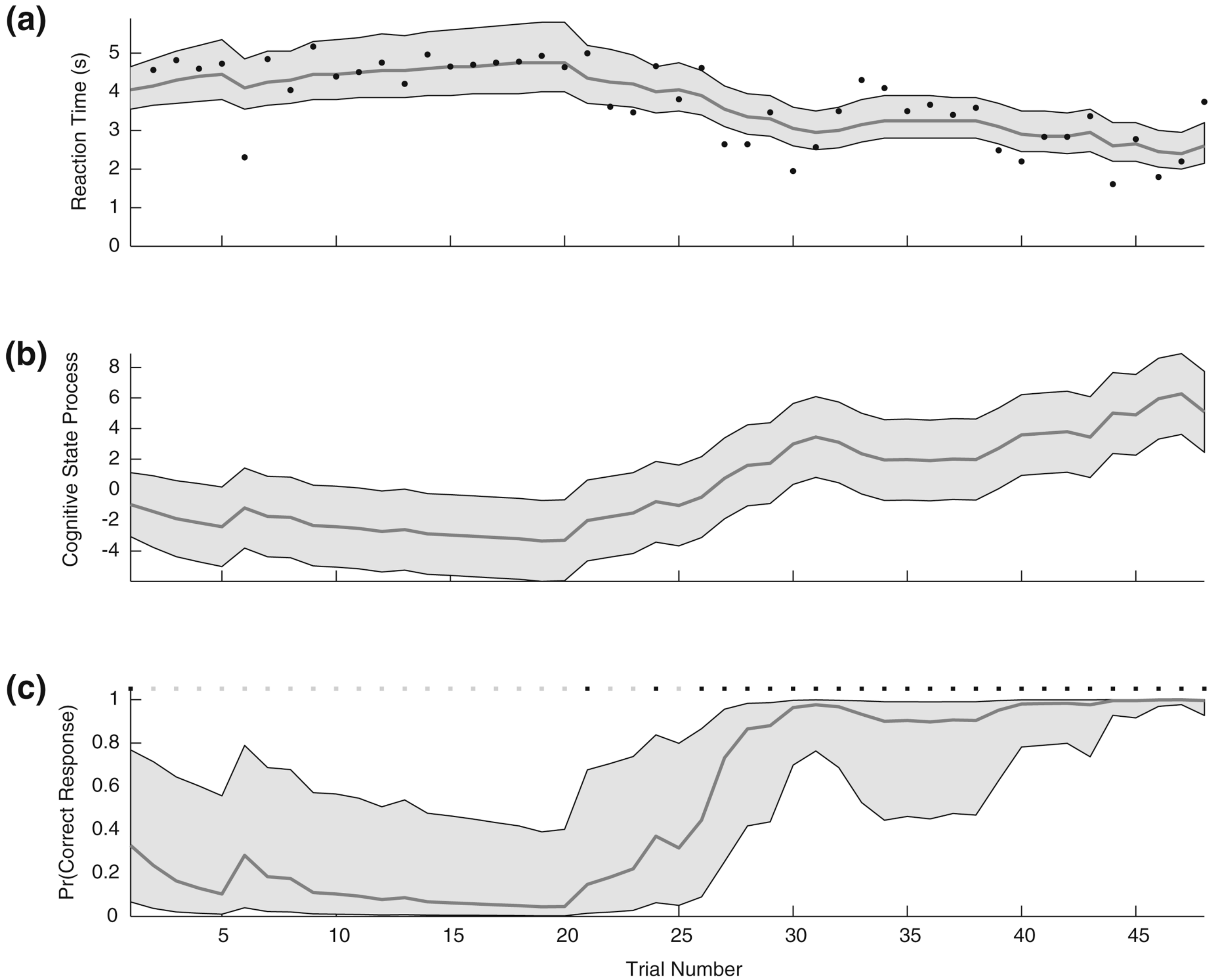

Fig. 6.

Mixed filter analysis of the reaction time and binary responses from the location-scene association experiment. **a** Mixed filter estimate of reaction time curve (*gray*). Black dots are the reaction times. **b** Mixed filter estimate of the cognitive state process (*gray*). **c** Mixed filter estimate of the learning curve (*gray*). Correct (*black*) and incorrect (*gray*) trial responses of the monkey are shown above **c**. The light gray regions are the 95% confidence intervals in each panel

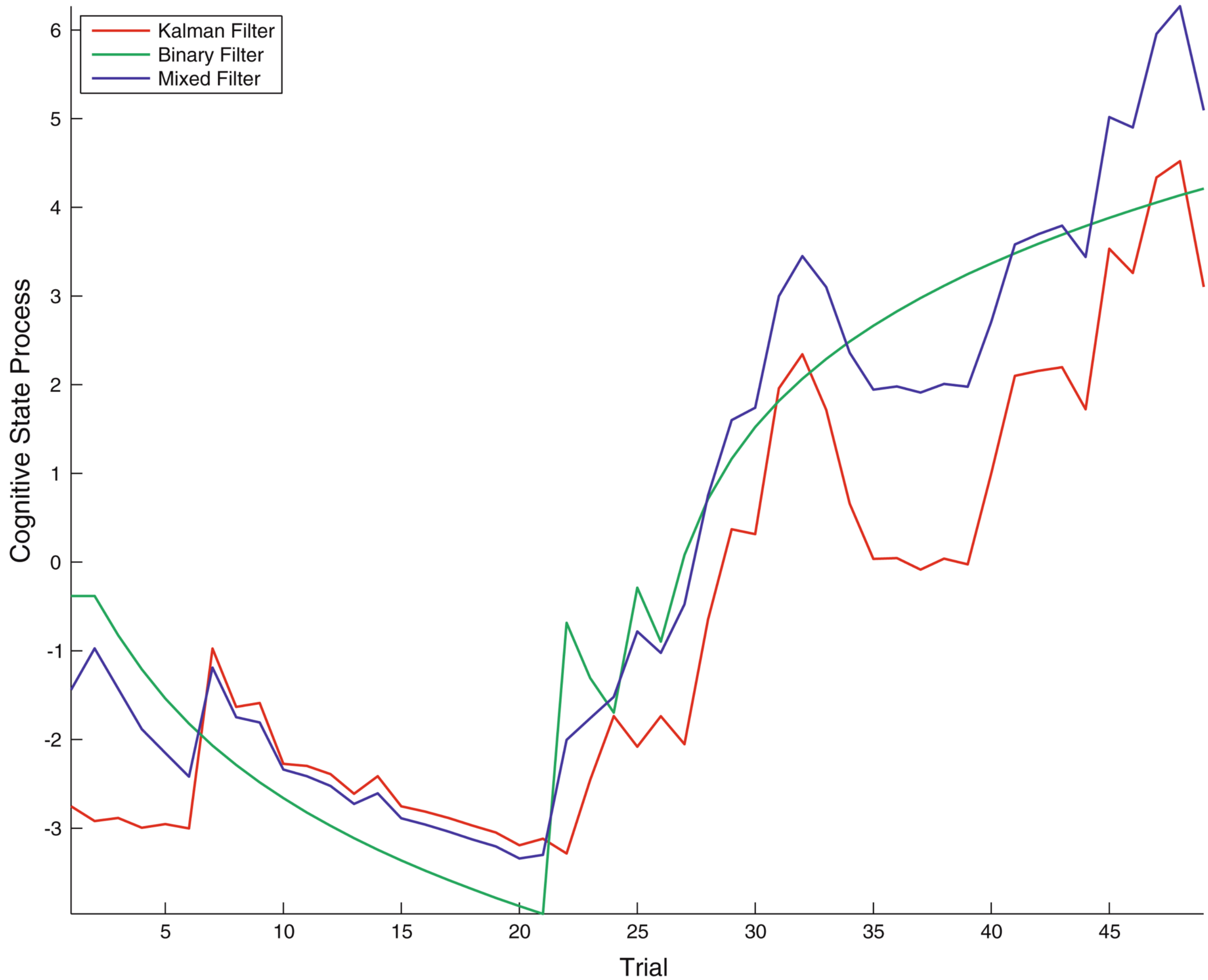

Fig. 7.
Binary filter (*green*), Kalman filter (*red*) and mixed filter (*blue*) estimates of the cognitive state process

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Mean-squared error, coverage probability for the 95% confidence intervals and median width of the 95% confidence intervals from the mixed, Kalman, and binary filter algorithms computed from 1,000 pairs of simulated binary and continuous performance measures from the learning experiments in Fig. 2

	Mixed filter	**Kalman filter**	**Binary filter**
Mean-squared error	0.6001	0.9132	1.0330
Coverage probability of 95% confidence intervals	0.9450	0.9150	0.9650
Median of 95% confidence interval widths	3.0063	3.3830	3.9398