Figure 3. Log Odds and Fisher Information as a Function of Time

The origin ($t = 0$) on all plots corresponds to the start of the integration of evidence, which about 220 ms after stimulus onset in the experimental data. (A–C) model; (D) data.

(A) Log odds for a binary decision as a function of time for four different levels of coherence (solid lines). Blue and black dotted lines: the coherence increases to 51.2% at $t = 100$ ms. After 100 ms, the slope matches the 51.2% coherence trials, as expected if the model is Bayes optimal.

(B) Fisher information as a function of time for continuous decision making (as in Figure 1B). Fisher information rises linearly with time, with higher slopes for higher coherences, as expected for Bayesian optimality. Dotted line: trial in which the coherence increases from 25.6% to 51.2%. In both (A) and (B), the kink at $t = 50$ ms is due to the discretization of time.

(C) Squares: Fisher information estimated by a single local linear estimator across all times and all coherences. Circles: Fisher information estimated by a local optimal estimator trained separately for each time and each coherence. Dotted lines: for each coherence, the upper line corresponds to the information estimated from the training set, while the lower trace is the information obtained from the testing set. The solid line is the average of the upper and lower dotted lines. The fact that both estimators return similar values of Fisher information shows that decoding LIP can be done nearly optimally without any knowledge of time or coherence. Green line: trials in which the coherence starts at 25.6% and then switches to 51.2% at 100 ms.

(D) Same as in (A) but for actual LIP neurons (n = 45; data from Roitman and Shadlen, 2002). The results are quantitatively similar to the model. The y axis is arbitrary up to a multiplicative factor and a DC offset.

posterior distributions encoded by this activity. As evidence accumulates in favor of 180°, the activity at 180° increases and the probability distribution becomes narrower. To determine whether this accumulation process is optimal, we can run the same test as in the binary case, except this time we use the average of the inverse of the variance of the posterior distributions (the Fisher information [Papoulis, 1991]) rather than the log odds. Figure 3B shows that, indeed, Fisher information increases linearly with time and the slope is an increasing function of coherence. Furthermore, when the coherence increases during the trial, so does the slope.

We now turn to the second question: can the maximum-likelihood estimate be computed from LIP activity, for all coherences and at all times, with a single attractor network? Because attractor networks are mathematically equivalent to local linear estimators (Deneve et al., 1999), this question can be rephrased as: is the performance of a single local linear estimator similar to the performances of a family of estimators, each specialized for one time and one coherence? Figure 3C shows that the Fisher information recovered by the specialized linear estimators is indeed very similar to the information recovered by a single one, hence demonstrating that a single attractor network can optimally decode LIP for all coherences and at all times.

Finally, we performed another test, now at decision time. With our framework, the network encodes a probability distribution at all times and in particular at decision time. This distribution reflects the quality of the data that have been accumulated and, consequently, the performance of the animal. Hence, for

both two- and four-choice experiments, the log odds estimated in the LIP layer should be higher at high coherence than at low coherence, since the performance of the animal is better in the former case. Figures 5A and 5B show that our model behaves as predicted. Note the important distinction with single-race bounded accumulation models (Bogacz et al., 2006; Huk and Shadlen, 2005; Link, 1992; Link and Heath, 1975; Ratcliff and Rouder, 1998). In such models, the state of the system is characterized by the value of the accumulation process. When the bound is hit, this value is always the same (Gold and Shadlen, 2001; Link, 1992; Shadlen et al., 2006a). Thus, there is no principled way to recover the probability that the decision is correct. An ad hoc solution has been proposed for two race models (Vickers, 1979), but it was not derived from probabilistic principles, and does not readily generalize to more than two choices.

Speed-Accuracy Tradeoff

When monkeys are tested on our decision making task in which they are free to choose when to respond, their psychometric and chronometric functions follow the profiles shown in Figures 6A and 6B. To obtain these curves with our model, we used a stopping rule similar to the one used in most models: a fixed bound on the maximum activity in the network (see Experimental Procedures). As can be seen, our model readily captures the performance and reaction time reported in monkeys whether the task involves two or four choices (data from Churchland et al., 2008). Moreover, the rate at which activity builds up on average