Temporal context calibrates interval timing

Mehrdad Jazayeri[1,2] & Michael N Shadlen[2]

We use our sense of time to identify temporal relationships between events and to anticipate actions. The degree to which we can exploit temporal contingencies depends on the variability of our measurements of time. We asked humans to reproduce time intervals drawn from different underlying distributions. As expected, production times were more variable for longer intervals. However, production times exhibited a systematic regression toward the mean. Consequently, estimates for a sample interval differed depending on the distribution from which it was drawn. A performance-optimizing Bayesian model that takes the underlying distribution of samples into account provided an accurate description of subjects' performance, variability and bias. This finding suggests that the CNS incorporates knowledge about temporal uncertainty to adapt internal timing mechanisms to the temporal statistics of the environment.

From simple habitual responses to complex sensorimotor skills, our behavioral repertoire exhibits a marked sensitivity to timing information. To internalize temporal contingencies and to put them to use in the control of conditioned and deliberative behavior, our nervous systems must be equipped with central mechanisms for processing time.

Among the elementary aspects of temporal processing, and one that has been the focus of many psychophysical studies of time perception, is the ability to measure the duration between events (that is, interval timing)[1]. A common feature associated with repeated estimation (or production) of a sample interval is that the s.d. of the estimated (or produced) intervals increases linearly with their mean, a property that is termed scalar variability[2–4]. Although previous work has shown how suitable forms of internal noise might lead to scalar variability[5–8], we do not know whether and how the nervous system can make use of this lawful relationship to improve timing behavior.

Scalar variability implies that measurements of relatively longer intervals are less reliable and thus more uncertain. We asked whether subjects have knowledge about this uncertainty and how they might exploit it to improve estimation and production of time intervals. This question is particularly important when one has prior expectations of how long an event might last. For example, if one measures an interval to be ~1.5 s, but expects it to be closer to 1.2 s on the basis of past experience, then they may conclude that the true interval was probably somewhere between 1.2 and 1.5 s. More generally, knowledge about the distribution of time intervals one may encounter, which we refer to as temporal context, could help to reduce uncertainty. The extent to which temporal context should inform temporal judgments depends on how unreliable measurements of time are. Although a metronome need not rely on temporal context to stay on the beat, a piano player may well use the tempo of a musical piece to coordinate finger movements in time. Thus, to make use of the oft-present temporal context, the brain must have knowledge about the reliability of its own measurements of time.

The question of how knowledge about temporal context may improve measurements of elapsed time can be posed rigorously in the framework of statistical inference. In this framework, to estimate a sample interval, the observer may take advantage of two sources of information: the likelihood function, which quantifies the statistics of sample intervals consistent with a measurement, and the prior probability distribution function of the sample intervals that the observer may encounter. One possibility is for the observer to ignore the prior distribution and to choose the most likely value directly from the likelihood function, a strategy known as the maximum-likelihood estimation (MLE)[9]. Alternatively, a Bayesian observer would combine the likelihood function and the prior and use some statistic to map the resulting posterior probability distribution onto an estimate. Common mapping rules are the maximum a posteriori (MAP) and Bayes least-squares (BLS), which correspond to the mode and the mean of the posterior, respectively.

To understand how humans evaluate their measurements of elapsed time in the presence of a temporal context, we asked subjects to estimate and subsequently reproduce time intervals in the subsecond-to-second range that were drawn from three different prior distributions. Subjects' production times showed a clear dependence on both the sample intervals and the prior distribution from which they were drawn. We fitted subjects' responses to various observer models, such as MLE, MAP and BLS, and found that a Bayesian observer associated with the BLS could account for the bias, variability and overall performance of every subject in all three prior conditions. This suggests that subjects have implicit knowledge of the reliability of their measurements of time and can use this information to adjust their timing behavior to the temporal regularities of the environment. Furthermore, our observer model indicated that this sophisticated Bayesian behavior can be accounted for by a nonlinear transformation that simply and directly maps noisy measurement of time to optimal estimates.

Figure 1 The ready-set-go time-reproduction task. (**a**) Sequence of events during a trial. Appearance of a central spot indicated the start of the trial. Subjects were instructed to fixate the central spot and maintain fixation throughout the trial. A white feedback spot was visible to the left of the fixation point. After a random delay (0.25–0.85 s), two briefly flashed cues, 'Ready' and 'Set', were presented in sequence. Subjects were instructed to estimate the sample interval demarcated by the time between the ready and set cues and to reproduce it immediately afterwards. The production times were measured from the time of the set cue to the time subjects responded via a key-press. When production times were in an experimentally adjusted window around the target interval (Online Methods), the feedback spot turned green to provide positive feedback. (**b**) Distribution of sample intervals. In each session, sample intervals were drawn randomly from one of three partially overlapping discrete uniform prior distributions (that is, short, intermediate and long) shown by the black, dark red and light red bar charts. (**c**) Feedback schedule. The

width of the window for which production times were positively reinforced (green area) scaled with the sample interval (Online Methods). No feedback was provided for early and late responses. The plot shows an example schedule for the intermediate prior condition.
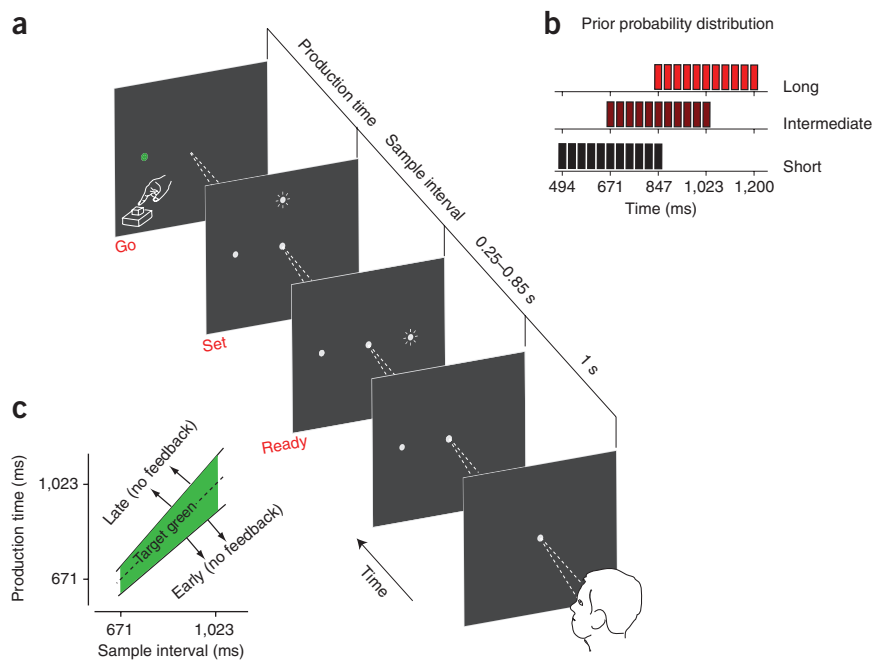
RESULTS

The ready-set-go procedure

Subjects had to measure and immediately afterwards reproduce different sample intervals. A sample interval, t_s, was demarcated by two brief flashes, a 'ready' cue followed by a 'set' cue. The corresponding production time, t_p, was measured from the time of the set cue to when subjects proactively responded via a manual key press (**Fig. 1a**). In each session, sample intervals were drawn from a discrete uniform prior distribution. For each subject, three partially overlapping prior distributions (short, intermediate and long) were tested (**Fig. 1b**). Subjects received feedback for sufficiently accurate production times (**Fig. 1c**). The main data for each prior condition were collected after an initial learning stage (typically 500 trials) to ensure that subjects had time to adapt their responses to the range of sample intervals presented.

Subjects' timing behavior exhibited three characteristic features (**Fig. 2**). First, production times monotonically increased with sample

intervals. Second, for each prior condition, production times were systematically biased toward the mean of the prior, as evident from their tendency to deviate from sample intervals and gravitate toward the mean interval[10–12]. Consequently, mean production times associated with a particular t_s were differentially biased for the three prior conditions. Third, production time biases were more pronounced in the intermediate and even more so in the long prior conditions, indicating that longer sample intervals were associated with progressively stronger prior-dependent biases. Similarly, in each prior condition, the magnitude of the bias was larger for the longest sample interval compared with the shortest sample interval (**Supplementary Fig. 1**).

Scalar variability implies that the measurement of longer sample intervals engender more uncertainty. According to Bayesian theory, for these more uncertain measurements, subjects should rely more on their prior expectation (**Supplementary Fig. 2**). This is consistent with the observed increases in prior-dependent biases associated with longer sample intervals and suggests that subjects might have adopted a Bayesian strategy to reproduce time intervals. We developed probabilistic observer models to evaluate these observations quantitatively and to understand the computations from which they might arise.

The observer model

The observer model's task was to reproduce the sample interval, t_s. As a result of measurement noise, the measured interval, t_m, may differ from t_s. The observer must use t_m to compute an estimate, t_e, for t_s. To do so, the observer may use an estimator that relies on probabilistic
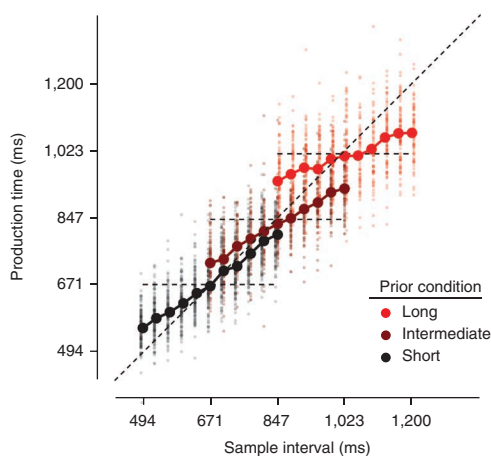

Figure 2 Time reproduction in different temporal contexts. Individual production times for every trial (small dots), and their averages for each sample interval (large circles connected with thick lines) are shown for three prior conditions for a typical subject. Average production times deviated from the line of equality (diagonal dashed line) toward the mean of the priors (horizontal dashed lines). Prior-dependent biases were strongest for the long prior condition. Color conventions are as described in **Figure 1b**.
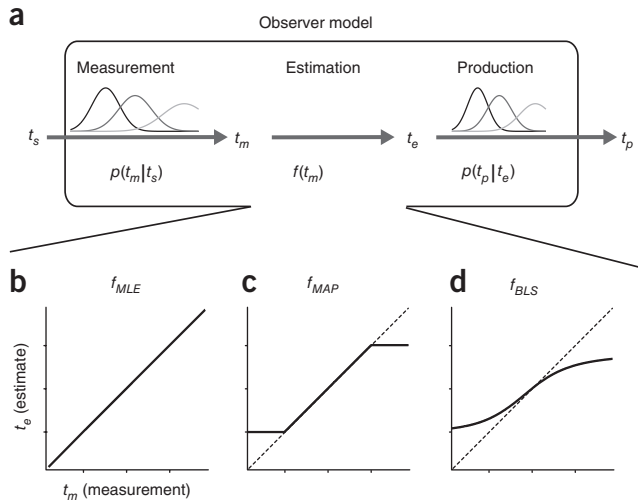
Figure 3 The observer model for time reproduction. (**a**) The three-stage architecture of the model. In the first stage, the sample interval, t_s, is measured. The relationship between the measured interval, t_m, and t_s is characterized by measurement noise, $p(t_m|t_s)$, which was modeled as a Gaussian function centered on t_s whose s.d. grows linearly with the mean (that is, scalar variability). The second stage is the estimator; that is, the deterministic function, $f(t_m)$, that maps t_m to t_e. The third stage uses t_e to produce interval t_p. The conditional dependence of t_p on t_e, $p(t_p|t_e)$, was characterized by production noise, which was modeled by a zero-mean Gaussian distribution whose s.d. scales linearly with t_e. (**b**–**d**) The deterministic mapping functions associated with the MLE, MAP and BLS models, respectively.

sources of information, such as the likelihood function and the prior distribution. However, the estimator itself is fully characterized by a deterministic function, f, that maps t_m to t_e; that is, $t_e = f(t_m)$. Finally, additional noise during the production phase may cause the production time, t_p, to differ from t_e (**Fig. 3a**).

To formulate the model mathematically, we need to specify the relationship between t_s, t_m, t_e, and t_p. The relationship between t_m and t_s can be quantified by the conditional probability distribution, $p(t_m|t_s)$, the probability of different measurements for a specific sample interval. This distribution also specifies the likelihood function, $\lambda_{t_m}(t_s)$, a statistical description of the different sample intervals associated with a fixed measurement. We modeled $p(t_m|t_s)$ as a Gaussian distribution centered at t_s and assumed that its s.d. grows linearly with its mean (**Fig. 3a**). This assumption was motivated by the scalar variability of timing. The distribution of measurement noise was thus fully characterized by the ratio of the s.d. to the mean of $p(t_m|t_s)$, which we will refer to as the Weber fraction associated with the measurement, w_m. With the same arguments in mind, we assumed that the distribution of t_p conditioned on t_e, $p(t_p|t_e)$, is also Gaussian, centered at t_e and associated with a constant Weber fraction, w_p.

Finally, the relationship between t_m and t_e was modeled by a deterministic mapping function, f, which we will refer to as the estimator. Different estimators are associated with different mapping rules. Among them, we focused on the MLE, MAP and BLS because of their well-known properties and because they were most germane to the development of our arguments with respect to the psychophysical data. We denote the corresponding estimators by f_{MLE}, f_{MAP} and f_{BLS}, respectively (**Fig. 3b**–**d**).

The f_{MLE} estimator assigns t_e to the peak of the likelihood function (**Fig. 4a**). In our model, with a Gaussian-distributed measurement noise and a constant Weber fraction, t_e would be proportional to t_m (Online Methods). The f_{MAP} and f_{BLS} estimators, on the other hand, rely on the posterior distribution, which is proportional to the product of the prior distribution and the likelihood function. Because the prior distribution that we used was uniform, the posterior was a scaled replica of the likelihood function in the domain of the prior and zero elsewhere. The MAP rule extracts the mode of the posterior, which would correspond to the peak of the likelihood function, except when the peak falls below or above the prior distribution's shortest or longest sample interval. Thus, f_{MAP} is the same as f_{MLE}, with the difference that its range is limited to the domain of the prior (**Fig. 4b**). For BLS, which is associated with the mean of the posterior, the estimator, f_{BLS}, is a sigmoid function of t_m (**Fig. 4c**). Note that as the specification of these estimators does not invoke any additional free parameters, the observer model associated with each estimator was fully characterized by just two free parameters, w_m and w_p.

Figure 4 MLE, MAP and BLS estimators. (**a**–**c**) Schematic representations of how MLE, MAP and BLS estimates are computed, respectively. Upward arrows in black and gray show two example sample intervals. Vertical dashed lines represent the noise-perturbed measurements associated with those sample intervals. Measured intervals differ from the corresponding samples as shown by the misalignment between the upward arrows and their corresponding vertical dashed lines. The likelihood functions associated with the two measurements are shown on the far right of each panel (rotated 90 degrees). These likelihood functions are plotted with respect to the measurements, as shown by the reflection of the measured interval on the diagonal (horizontal dashed lines). The MLE estimator is shown in **a**. The peak of the likelihood function determines the estimate (filled circles, thick left arrow). The corresponding mapping function, f_{MLE}, for all possible measurements is shown by the solid black line with the two example cases superimposed (Online Methods). The MAP estimator is shown in **b**. Right, the posterior distributions (truncated Gaussian functions) for the two measurements are computed by multiplying their associated likelihood functions by the prior (gray bar chart). MAP estimates are computed from the mode of the posterior (filled circles). The corresponding mapping function, f_{MAP}, is the same as f_{MLE} with the difference that its range is limited by the domain of the prior. The BLS estimator is shown in **c**. Data are presented as in **b**, except that for BLS, the mean of the posterior determines the estimate. The resulting mapping function, f_{BLS}, is sigmoidal in shape.
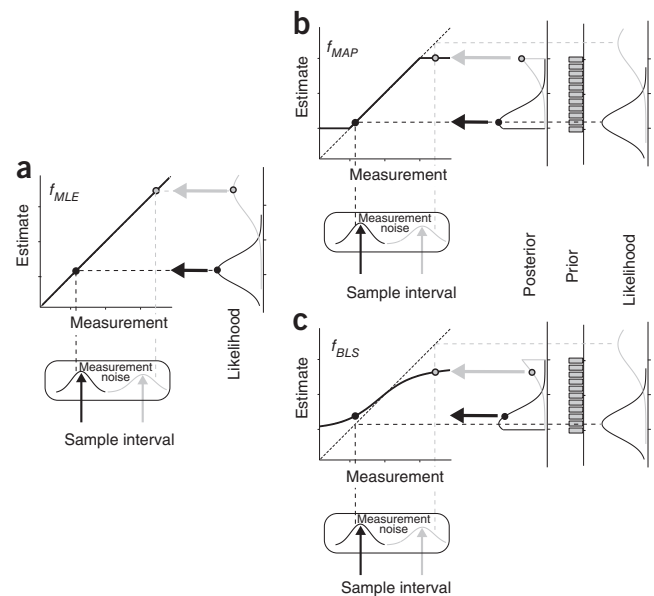
Figure 5 Time-reproduction behavior in humans and model observers.
(**a**) For each sample interval (referred to by subscript i) in each prior condition, we computed two statistics: $BIAS_i$ and VAR_i. $BIAS_i$ is the average difference between the production times and the sample interval and VAR_i as the corresponding variance. As an example, the plots shows how the $BIAS_i$ and VAR_i were computed for the largest sample interval associated with the long prior condition for one subject (data are presented as in **Fig. 2**). For this distribution of production times (histogram), $BIAS_i$ is the difference between the solid horizontal red line and the horizontal dashed line and VAR_i is the corresponding variance. For each prior condition, we computed two summary statistics; $BIAS$ is the root mean square of $BIAS_i$ and VAR is the average of VAR_i across sample intervals. (**b**) $\sqrt{VAR}$ versus $BIAS$ for three prior conditions for the same subject as in **a**. On a plot of $\sqrt{VAR}$ against $BIAS$, the locus of a constant RMSE value is a quarter circle. Dashed quarter circles show the loci of RMSE values associated with the $\sqrt{VAR}$ and $BIAS$ derived from the subject's production times. (**c**) Simulated production times from the best-fitted MLE model to the data in **a**. (**d**) The scatter of $\sqrt{VAR}$ and $BIAS$ of the best-fitted MLE model for three prior conditions (small dots) computed from 100 simulations similar to the one shown in **c**. The $\sqrt{VAR}$ and $BIAS$ of the subject are plotted for comparison (same as in **b**). (**e**–**h**) Data are presented as in **c** and **d** and show results for the best-fitted MAP (**e**,**f**) and BLS (**g**,**h**) models, respectively. Color conventions are as described in **Figure 1b**.

### Comparing experimental data with the observer model

Our psychophysical data consisted of pairs of sample intervals and production times (t_s and t_p), but the observer model that we created to relate t_s to t_p relies on two intervening and unobservable (hidden) variables, t_m and t_e. We thus expressed these two hidden variables in terms of their probabilistic relationship to the observable variables t_s and t_p (Online Methods) and derived a direct relationship between production times and sample intervals. This formulation was then used to examine which of the three observer models described human subjects' responses best.

To compare human subjects' responses to those predicted by the observer models, we quantified production times with two statistics, their variance (VAR) and bias ($BIAS$) (**Fig. 5a**), which together partition the overall root mean squared error ($RMSE$) by $RMSE^2=VAR+BIAS^2$. This relationship, which highlights the familiar trade-off between the VAR and BIAS, when written as the sum of squares, becomes the standard equation of a circle, $RMSE^2 = \left(\sqrt{VAR}\right)^2 + BIAS^2$.

This geometric description indicates that in a plot of $\sqrt{VAR}$ versus $BIAS$, a continuum of values along a quarter circle would lead to the same $RMSE$ (**Fig. 5b**). It also provides a convenient graphical description for how a larger $RMSE$ represented by a quarter circle with a larger radius may arise from increases in $\sqrt{VAR}$, $BIAS$ or both. We used this plot to summarize the statistics of production times and to evaluate the degree to which different observer models could capture those statistics.

We fitted the parameters of the MLE, MAP and BLS models (w_m and w_p) for each subject on the basis of the production times in the three prior conditions. We then simulated each subject's behavior using the fitted observer models and compared each model's predictions to the actual responses using the $BIAS$, $\sqrt{VAR}$ and $RMSE$ statistics (**Fig. 5c**–**g**).

The MLE model did not exhibit the prior-dependent biases present in production times (**Fig. 5c**,**d**), because it does not take the prior into account. This failure cannot be attributed to an unsuccessful fitting procedure or a misrepresentation of the likelihood function. The fact that subjects' production times depended on the prior condition would render any estimator that neglects the prior inadequate, the parametric form of the likelihood function
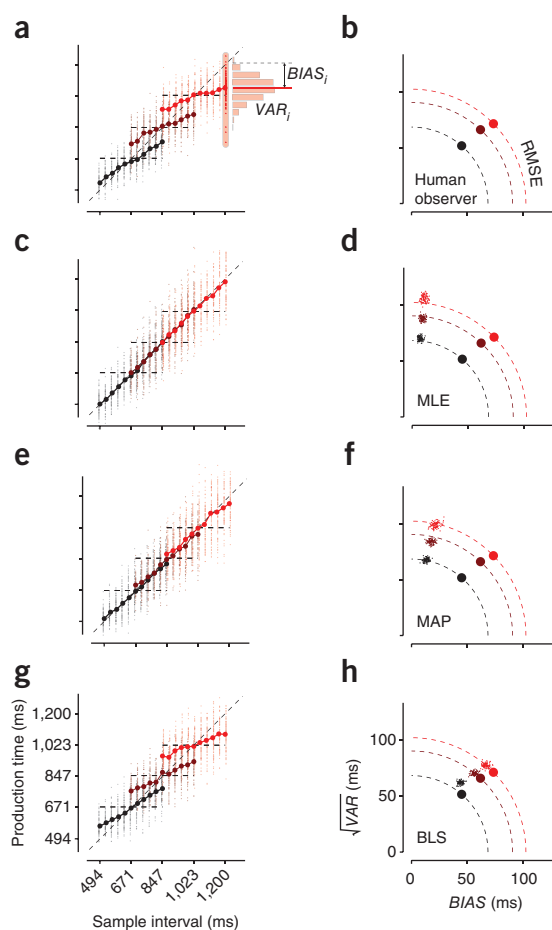
notwithstanding. The MAP model was slightly better than the MLE model at capturing the trade-off between BIAS and VAR (**Fig. 5e**,**f**), but it also underestimated the bias of the production times and overestimated their variance for all subjects. The BLS model on the other hand, mimicked the bias and variance of the production times quite well (**Fig. 5g**). It captured the overall RMSE, as well as the trade-off between the VAR and the BIAS (**Fig. 5h**), and was statistically superior to both MLE and MAP estimators across our subjects (**Fig. 6**).

We evaluated several variants of the BLS model by incorporating different assumptions concerning the measurement and production noise. In our main model (**Fig. 4c**), we fit Weber fractions for both sources of noise (w_m and w_p), consistent with the observation that, for all subjects, the s.d. of the production times was roughly proportional to the mean (**Supplementary Fig. 3**). We also considered the possibility that the s.d. of either the measurement noise or the production noise scales with the base interval, whereas the other noise source has constant s.d. (**Supplementary Tables 1** and **2**). For all subjects, the original BLS model outperformed the model in which the measurement noise had a constant s.d. and, for five out of six subjects, it outperformed the alternative in which the production noise had a constant s.d. (Akaike Information Criterion; **Supplementary Table 1**). Moreover, a BLS model in which Weber fractions were assumed to be identical ($w_m = w_p$) was inferior to the original BLS model (log likelihood ratio test for nested models, $P < 0.03$ for one subject and $P < 10^{-7}$ for others). The importance of the measurement and production Weber fractions in accounting for the bias and variability of production times was also evident in model simulations (**Supplementary Fig. 4**).
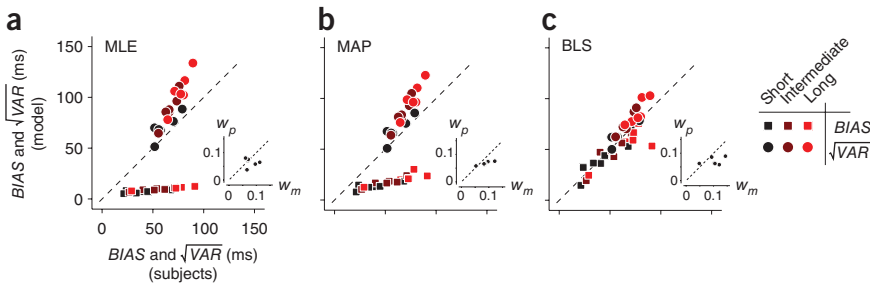
Figure 6 Time reproduction behavior in humans and model observers: model comparison. (**a**) Average *BIAS* (squares) and $\sqrt{VAR}$ (circles) computed from 100 simulations of the best-fitted MLE model as a function of *BIAS* and $\sqrt{VAR}$ computed directly from psychophysical data for all six subjects and all three prior conditions. The inset shows the Weber fraction of the measurement and production noise (w_p versus w_m) of the best-fitted MLE model for the six subjects. (**b,c**) Data are presented as in **a** for the MAP and BLS models, respectively. Each subject contributed six data points to each panel; that is, three prior conditions (black, dark red and light red) by two metrics (*BIAS* and $\sqrt{VAR}$).

Because our observer models were described by just two parameters (w_m and w_p) and all of the models used the same number of parameters, we were reasonably confident that the success of the BLS rule was not a result of over-fitting. Nonetheless, we tested for this possibility by fitting the model to data from the short and long prior conditions. The fits captured the statistics of the intermediate prior condition equally well. Finally, we note that the fits for the BLS and MAP rules did not differ systematically (**Fig. 6a–c**). Therefore, the success of the BLS model cannot be attributed to the constraints inherent in our fitting procedure, but rather to its superior description of the estimator subjects adopted in this task.

DISCUSSION
Our central finding is that humans can exploit the uncertainty associated with measurements of elapsed time to optimize their timed responses to the statistics of the intervals that they encounter. This conclusion is based on the success of a Bayesian observer model that accurately captured the statistics of subjects' production times in a simple time-reproduction task.

A characteristic feature of subjects' production times was that they were systematically biased toward the mean of the distribution of sample intervals. This observation is consistent with the ubiquitous central tendency of psychophysical responses in categorical judgment and motor production[10–14]. Previous work, such as the adaptation-level theory[14] and range-frequency theory[13], attributed these so-called range effects to subjects' tendency to evaluate a stimulus on the basis of its relation to the set of stimuli from which it is drawn. These theories, however, do not offer an explanation for what gives rise to such range effects in the first place and whether they are of any value. In contrast, our work suggests that it is subjects' (implicit) knowledge of their temporal uncertainty that determines the strength of the range effect. Moreover, the Bayesian account of range effects suggest that production time biases help, rather than harm, subjects' overall performance (**Supplementary Figs. 2** and **5**).

Bayesian interval timing
Bayesian models have had great success in describing a variety of phenomena in vision and sensorimotor control[15–18] and interval timing[19,20]. Symptomatic to these models are prior-dependent biases whose magnitude increases for progressively less reliable measurements[21]. Motivated by the observation of such biases in our subjects' behavior and the success of a previous Bayesian model of coincidence

timing[19], we set out to formulate a Bayesian model for time reproduction.

The model consisted of three stages. The first stage emulated a noisy measurement process that quantified the probabilistic relationship between the sample intervals and the corresponding noise-perturbed measurements[22]. In the second stage, a Bayesian estimator computed an estimate of the sample interval from the measurement. Finally, a noisy production stage converted estimates to production times[23,24]. Consistent with previous work on interval timing, the measurement and production noise exhibited scalar variability[2,3,5,7].

The estimator in the second stage of the model defines a deterministic mapping of measurements to estimates and its functional form is determined precisely from the likelihood function, the prior distribution and the cost (loss) function. The success of a Bayesian estimator therefore depends on how well the likelihood, the prior and the cost function are constrained.

In psychophysical settings, as sensory measurements are not directly accessible, the likelihood function must be inferred from behavior and suitable assumptions about the distribution of noise. For example, cue-combination studies make the reasonable assumption that measurements are perturbed by additive zero-mean Gaussian noise and infer the width of the likelihood function from psychophysical thresholds[25,26]. Alternatively, it is possible to model the likelihood function on the basis of the uncertainty associated with external noise in the stimulus[16,27,28]. We modeled the likelihood on the basis of the assumption that the distribution of measurements associated with a sample interval was Gaussian, centered on the sample interval and had a s.d. that scaled with the mean (Online Methods).

To tease apart the roles of the likelihood function and the prior, it is important to be able to vary them independently. One common strategy for manipulating likelihoods is to control the factors that change psychophysical thresholds, such as varying the external noise in the stimulus[16,27]. We exploited the scalar variability of timing to manipulate likelihoods. This property, which arises from internal noise only and is known to hold across tasks and species[2–4] for the range of times that we used[10], allowed us to manipulate the likelihood function simply by changing the sample interval. To manipulate the prior independently, we collected data using three discrete uniform prior distributions. The priors were partially overlapping so that certain sample intervals were tested for two or three different priors, which enabled us to evaluate the effect of the prior independent of the likelihood function.

To convert the posterior distribution to an estimate, we needed to specify the cost function associated with the estimator. We considered two possibilities: a cost function that penalizes all erroneous estimates similarly, which corresponds to the mode of the posterior (MAP), and a cost function that penalizes errors by the square of their magnitude, which corresponds to the mean of the posterior (BLS). We also considered a maximum-likelihood estimator that ignores the prior and chooses the peak of the likelihood function for the estimate (MLE). To decide which of these estimators better described subjects' behavior, it was essential to consider both the bias and the variability of production times. This technique, which was originally introduced to estimate internal priors from psychophysical data[22], provided a powerful constraint in the specification of the estimator's mapping function.

We used our three-stage model to estimate the measurement and production Weber fractions, and to decide which of the three mapping rules (MLE, MAP or BLS) better captured production times[29]. The MLE estimator clearly failed to capture the pattern of prior-dependent biases evident in every subject's production times, as expected from any estimator that neglects the prior. By incorporating the prior, both the MAP and BLS estimators exhibited contextual biases, but the BLS consistently outperformed the MAP model in explaining the trade-off between the trial-to-trial variability and bias across our subjects (**Fig. 6b,c**). It is important to emphasize that, had we ignored the trial-to-trial variability, both BLS and MAP, as well as a variety of other Bayesian models, could have accounted for the prior-dependent biases in our data.

We also considered variants of the BLS model in which either the measurement or production noise were modeled as Gaussian with a fixed s.d. (not scalar). Overall, our original model outperformed these alternatives (**Supplementary Table 1**), as the measurement and production Weber fractions had different effects on the bias and variance of production times (**Supplementary Fig. 4**). The degrading effect of formulating noise with a fixed s.d. was more severe for the measurement stage than it was for the production stage (**Supplementary Table 1**).

Despite the success of our modeling exercise, further validation is required to substantiate the role of a BLS mapping in interval timing. Four considerations deserve scrutiny. First, formulation of the likelihood function might take into account factors other than scalar variability that could alter measurement noise. For example, task difficulty or reinforcement schedule (**Supplementary Fig. 6**) could motivate subjects to pay more attention to certain intervals and to measure them more reliably, which could in turn strengthen the role of the likelihood function relative to the prior. Therefore, it is important to consider attention and other related cognitive factors as an integral part of how the nervous system could balance the relative effects of the likelihood function and the prior. Second, knowledge of the prior is itself subject to uncertainty and the internalized prior distribution may differ from the one imposed experimentally. Third, the feedback subjects receive probably interacts with the mapping rule that they adopt. Our feedback schedule did not encourage the use of a BLS rule, but we cannot rule out the possibility that it influenced subjects' behavior. Fourth, although the operation of a Bayesian estimator is formulated deterministically, its neural implementation is probably subject to biological noise. These different sources of variability must be parsed out before the estimator can be characterized definitively. These considerations, which concern all Bayesian models of psychophysical data, highlight the gap between normative descriptions and their biological implementation.

We referred to our model as a Bayesian observer and not a Bayesian observer-actor because our formulation was only concerned with making optimal estimates. However, as the full task of the observer was to reproduce those estimated intervals, we can formulate a Bayesian observer-actor whose objective is to directly optimize production times and not the intervening estimates. This model has to incorporate the measurement uncertainty, the production uncertainty and the prior probability distribution to compute the probability of every possible pair of sample and production interval. It would then use this joint posterior to minimize the cost of producing erroneous intervals. The derivations associated with the Bayesian observer-actor model are more involved and beyond the scope of our work. However, we note that under suitable assumptions, the two models would behave similarly (Online Methods).

Context-dependent central timing

Our findings suggest that the brain takes into account knowledge of temporal uncertainty and adapts its time keeping mechanisms to temporal statistics in the environment. What neural computations may lead to such sophisticated behavior? One possibility is that the brain implements a formal Bayesian algorithm. For example, populations of neurons might maintain an internal representation of the prior distribution and the likelihood function, multiply them to represent a posterior and produce an estimate by approximating its expectation. Related variants of this scheme are also conceivable. For instance, our results could be accommodated by an MLE strategy if the prior would exert its effect indirectly by changing the statistics of noise associated with measurements. Another, more attractive possibility that obviates the need for explicit representations of the likelihood function and the prior is for the brain to learn the sensorimotor transformation that would map measurements onto their corresponding Bayesian estimates directly[30]. This is what our observer model exemplifies; it establishes a deterministic nonlinear mapping function to directly transform measurements to estimates. Evidently, this form of learning must incorporate knowledge about scalar variability and prior distribution.

Electrophysiological recordings from sensorimotor structures in monkeys have described computations akin to those that our observer model utilizes. For instance, parietal association regions and subcortical neurons in caudate have been shown to reflect flexible sensorimotor associations[31,32]. The time course of activity across sensorimotor neurons is believed to represent sensory evidence[33], its integration with the prior information[34], and the preparatory signals in anticipation of instructed and self-generated action[35–37]. The importance of sensorimotor structures in time reproduction is further reinforced by their consistent activation in human neuroimaging studies that involve time sensitive computations[38–41].

A variety of models have been proposed to explain the perception and use of an interval of time. Information theoretic models attribute the sense of time to the accumulation of tics from a central clock[11,42,43]; physiological studies have noted a general role for rising neural activity for tracking elapsed time in the brain[36,37,44–48] and biophysical models have been developed that suggest that time may be represented through the dynamics of neuronal network[49]. Our work, which does not commit to a specific neural implementation, suggests that the internal sense of elapsed time in the subsecond-to-second range may arise from a plastic sensorimotor process that enables us to operate efficiently in different temporal contexts.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

AUTHOR CONTRIBUTIONS
M.J. designed the experiment, collected and analyzed the data and performed the computational modeling. M.N.S. helped in data analysis and provided intellectual support throughout the study. M.J. and M.N.S. wrote the manuscript.

ARTICLES

1. Mauk, M.D. & Buonomano, D.V. The neural basis of temporal processing. *Annu. Rev. Neurosci.* **27**, 307–340 (2004).
2. Gallistel, C.R. & Gibbon, J. Time, rate, and conditioning. *Psychol. Rev.* **107**, 289–344 (2000).
3. Rakitin, B.C. *et al.* Scalar expectancy theory and peak-interval timing in humans. *J. Exp. Psychol. Anim. Behav. Process.* **24**, 15–33 (1998).
4. Brannon, E.M., Libertus, M.E., Meck, W.H. & Woldorff, M.G. Electrophysiological measures of time processing in infant and adult brains: Weber's Law holds. *J. Cogn. Neurosci.* **20**, 193–203 (2008).
5. Gibbon, J. & Church, R.M. Comparison of variance and covariance patterns in parallel and serial theories of timing. *J. Exp. Anal. Behav.* **57**, 393–406 (1992).
6. Reutimann, J., Yakovlev, V., Fusi, S. & Senn, W. Climbing neuronal activity as an event-based cortical representation of time. *J. Neurosci.* **24**, 3295–3303 (2004).
7. Matell, M.S. & Meck, W.H. Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Brain Res. Cogn. Brain Res.* **21**, 139–170 (2004).
8. Ahrens, M. & Sahani, M. Inferring elapsed time from stochastic neural processes. in *Advances in Neural Information Processing Systems* (eds. Platt, J.C. Koller, D. Singer, Y. & Roweis, S.) (MIT Press, Cambridge, Massachusetts, 2008).
9. Casella, G. & Berger, R.L. (Duxbury Resource Center, Pacific Grove, California, 2002).
10. Lewis, P.A. & Miall, R.C. The precision of temporal judgment: milliseconds, many minutes, and beyond. *Phil. Trans. R. Soc. Lond. B* **364**, 1897–1905 (2009).
11. Treisman, M. Temporal discrimination and the indifference interval. Implications for a model of the "internal clock". *Psychol. Monogr.* **77**, 1–31 (1963).
12. Hollingworth, H.L. The central tendency of judgement. *Arch. Psychol.* **4**, 44–52 (1913).
13. Parducci, A. Category judgment: a range-frequency model. *Psychol. Rev.* **72**, 407–418 (1965).
14. Helson, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.* **55**, 297–313 (1948).
15. Kersten, D., Mamassian, P. & Yuille, A. Object perception as Bayesian inference. *Annu. Rev. Psychol.* **55**, 271–304 (2004).
16. Körding, K.P. & Wolpert, D.M. Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247 (2004).
17. Knill, D.C. & Richards, W. *Perception as Bayesian Inference* (Cambridge University Press, Cambridge, 1996).
18. Mamassian, P., Landy, M.S. & Maloney, L.T. Bayesian modeling of visual perception. in *Probabilistic Models of the Brain: Perception and Neural Function* (eds. Rao, R.P.N., Olshausen, B.A. & Lewicki, M.S.) 239–286 (MIT Press, Cambridge, Massachusetts, 2002).
19. Miyazaki, M., Nozaki, D. & Nakajima, Y. Testing Bayesian models of human coincidence timing. *J. Neurophysiol.* **94**, 395–399 (2005).
20. Hudson, T.E., Maloney, L.T. & Landy, M.S. Optimal compensation for temporal uncertainty in movement planning. *PLOS Comput. Biol.* **4**, e1000130 (2008).
21. Bernardo, J.M. & Smith, A.F.M. *Bayesian Theory* (Wiley, New York, 1994).
22. Stocker, A.A. & Simoncelli, E.P. Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* **9**, 578–585 (2006).
23. Trommershäuser, J., Maloney, L.T. & Landy, M.S. Statistical decision theory and the selection of rapid, goal-directed movements. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **20**, 1419–1433 (2003).
24. Mamassian, P. Overconfidence in an objective anticipatory motor task. *Psychol. Sci.* **19**, 601–606 (2008).
25. Ernst, M.O. & Banks, M.S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
26. Jacobs, R.A. Optimal integration of texture and motion cues to depth. *Vision Res.* **39**, 3621–3629 (1999).
27. Tassinari, H., Hudson, T.E. & Landy, M.S. Combining priors and noisy visual cues in a rapid pointing task. *J. Neurosci.* **26**, 10154–10163 (2006).
28. Graf, E.W., Warren, P.A. & Maloney, L.T. Explicit estimation of visual uncertainty in human motion processing. *Vision Res.* **45**, 3050–3059 (2005).
29. Körding, K.P. & Wolpert, D.M. The loss function of sensorimotor learning. *Proc. Natl. Acad. Sci. USA* **101**, 9839–9842 (2004).
30. Raphan, M. & Simoncelli, E.P. Learning to be Bayesian without supervision. in *Neural Information Processing Systems* 1145–1152 (MIT Press, Cambridge, Massachusetts, 2006).
31. Toth, L.J. & Assad, J.A. Dynamic coding of behaviorally relevant stimuli in parietal cortex. *Nature* **415**, 165–168 (2002).
32. Lauwereyns, J. *et al.* Feature-based anticipation of cues that predict reward in monkey caudate nucleus. *Neuron* **33**, 463–473 (2002).
33. Shadlen, M.N. & Newsome, W.T. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* **86**, 1916–1936 (2001).
34. Gold, J.I., Law, C.T., Connolly, P. & Bennur, S. The relative influences of priors and sensory evidence on an oculomotor decision variable during perceptual learning. *J. Neurophysiol.* **100**, 2653–2668 (2008).
35. Janssen, P. & Shadlen, M.N. A representation of the hazard rate of elapsed time in macaque area LIP. *Nat. Neurosci.* **8**, 234–241 (2005).
36. Maimon, G. & Assad, J.A. A cognitive signal for the proactive timing of action in macaque LIP. *Nat. Neurosci.* **9**, 948–955 (2006).
37. Schultz, W. & Romo, R. Role of primate basal ganglia and frontal cortex in the internal generation of movements. I. Preparatory activity in the anterior striatum. *Exp. Brain Res.* **91**, 363–384 (1992).
38. Meck, W.H., Penney, T.B. & Pouthas, V. Cortico-striatal representation of time in animals and humans. *Curr. Opin. Neurobiol.* **18**, 145–152 (2008).
39. Cui, X., Stetson, C., Montague, P.R. & Eagleman, D.M. Ready...go: amplitude of the FMRI signal encodes expectation of cue arrival time. *PLoS Biol.* **7**, e1000167 (2009).
40. Nobre, A., Correa, A. & Coull, J. The hazards of time. *Curr. Opin. Neurobiol.* **17**, 465–470 (2007).
41. Rao, S.M., Mayer, A.R. & Harrington, D.L. The evolution of brain activation during temporal processing. *Nat. Neurosci.* **4**, 317–323 (2001).
42. Allan, L.G. Perception of time. *Percept. Psychophys.* **26**, 340–354 (1979).
43. Creelman, C.D. Human discrimination of auditory duration. *J. Acoust. Soc. Am.* **34**, 582–593 (1962).
44. Lee, I.H. & Assad, J.A. Putaminal activity for simple reactions or self-timed movements. *J. Neurophysiol.* **89**, 2528–2537 (2003).
45. Mita, A., Mushiake, H., Shima, K., Matsuzaka, Y. & Tanji, J. Interval time coding by neurons in the presupplementary and supplementary motor areas. *Nat. Neurosci.* **12**, 502–507 (2009).
46. Okano, K. & Tanji, J. Neuronal activities in the primate motor fields of the agranular frontal cortex preceding visually triggered and self-paced movement. *Exp. Brain Res.* **66**, 155–166 (1987).
47. Tanaka, M. Cognitive signals in the primate motor thalamus predict saccade timing. *J. Neurosci.* **27**, 12109–12118 (2007).
48. Tanaka, M. Inactivation of the central thalamus delays self-timed saccades. *Nat. Neurosci.* **9**, 20–22 (2006).
49. Buonomano, D.V. & Maass, W. State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* **10**, 113–125 (2009).

ONLINE METHODS

Psychophysical procedures. Six human subjects aged 19–40 years participated in this study after giving informed consent. All had normal or corrected-to-normal vision and all were naive to the purpose of the experiment. Subjects viewed all stimuli binocularly from a distance of 52 cm on a 17-inch iiyama AS4311U LCD monitor at a resolution of 1,024 × 768 driven by an Intel Macintosh G5 computer at a refresh rate of 85 Hz in a dark, quiet room.

In a ready-set-go time-reproduction task, subjects measured certain sample intervals demarcated by a pair of flashed stimuli and reproduced those intervals by producing time-sensitive manual responses. Each trial began with the presentation of a central fixation point for 1 s, followed by the presentation of a warning stimulus at a variable distance to the left of the fixation point. After a variable delay ranging from 0.25–0.85 s drawn randomly from a truncated exponential distribution, two 100-ms flashes separated by the sample interval, t_s, were presented. The first flash, which signified the ready stimulus, was presented at the same distance as the warning stimulus but to the right of the fixation point. The set stimulus was presented t_s ms afterwards and 5 degrees above the fixation point (**Fig. 1a**). Subjects were instructed to measure and reproduce the sample interval by pressing the space bar on the keyboard t_s ms after the presentation of the set. Production times, t_p, were measured from the center of the set flash (that is, 50 ms after its onset) to when the key was pressed. When t_p was sufficiently close to t_s, the warning stimulus changed from white to green to provide positive feedback and encourage stable performance.

All stimuli were circular in shape and were presented on a dark gray background. Except for the fixation point, which subtended 0.5 degrees of visual angle, all other stimuli were 1.5 degrees. To ensure that subjects could not use the layout of the stimuli to adopt a spatial strategy for the time-reproduction task (for example, track an imaginary moving target), we varied the distance of the ready and the warning stimulus from the fixation point on each trial (range of 7.5–12.5 degrees).

For each subject, three experimental conditions were tested separately. These conditions were the same in all respects except that, for each condition, the sample intervals were drawn from a different prior probability distribution. All priors were discrete uniform distributions with 11 values, ranging from 494–847 ms for the short, 671–1,023 ms for the intermediate and 847–1,200 ms for long prior condition. Note that to help tease apart the effects of prior condition from sample interval, the priors were chosen to be partially overlapping.

For each subject, the order in which the three prior conditions were tested was randomized. For each prior condition, subjects were tested after they completed an initial learning stage. Learning was considered to be complete when the variance and bias of the production times had stabilized (less than 10% change between sessions). The main data for each prior condition were collected in two sessions after learning for that condition was complete. Learning for each subsequent prior condition started after testing for the preceding prior condition was completed. For five of six subjects, the learning was completed by the end of the first session (less than 10% change between first and second sessions). For one subject, learning of the first prior condition was completed after four sessions. For this subject, the fifth and sixth sessions provided data for the first prior condition. For the other two prior conditions, similar to other subjects, responses stabilized after one practice session. All subjects typically participated in three sessions per week and each sessions lasted ~45 min (that is, nearly 500 trials).

Subjects received positive feedback for responses that fell in a specified window around t_s (that is, correct trials). To compensate for the increased difficulty associated with longer sample intervals, a natural consequence of scalar timing variability[2-4], the width of this window was scaled with the sample interval with a constant of proportionality, k. To ensure that the performance was comparable across different prior conditions, the value of k was controlled by an adaptive one-up, one-down procedure that added or subtracted 0.015 to or from k for each miss or correct trial. As such, every subject's performance for every session yielded approximately 50% positively reinforced trials (mean = 51.7%, s.d. = 1.33%). For each prior condition, the maximum (minimum) number of correct trials corresponded to the intermediate (extreme) sample intervals (**Supplementary Fig. 3**).

The Bayesian estimator. The noise distribution associated with the measurement stage of the model determines the distribution of t_m for a given t_s, $p(t_m|t_s)$. From the perspective of the observer who makes a measurement t_m, but does not know t_s, this relationship becomes a function of t_s that is known as the likelihood function, $\lambda_{t_m}(t_s) \equiv p(t_m|t_s)$, in which t_m is fixed. We modeled $p(t_m|t_s)$ as a Gaussian distribution with mean t_s and s.d. $w_m t_s$, which scaled linearly with t_s (scalar variability) with a constant coefficient of variation, w_m.

$$\lambda_{t_m}(t_s) \equiv p(t_m|t_s) = \frac{1}{\sqrt{2\pi(w_m t_s)^2}} e^{\frac{-(t_s - t_m)^2}{2(w_m t_s)^2}} \tag{1}$$

Similarly, the production noise distribution, $p(t_p|t_e)$, was assumed to be Gaussian with zero mean and a constant coefficient of variation, w_p.

$$p(t_p|t_e) = \frac{1}{\sqrt{2\pi(w_p t_e)^2}} e^{\frac{-(t_p - t_e)^2}{2(w_p t_e)^2}} \tag{2}$$

To simplify derivations, we modeled the discrete uniform prior distributions used in the experiment as continuous. For each prior condition, we specified the domain of sample intervals between $t_s^{\min}$ and $t_s^{\max}$ on the basis of the minimum and maximum values used in the experiment.

$$\pi(t_s) = \begin{cases} \dfrac{1}{t_s^{\max} - t_s^{\min}} & \text{for} \quad t_s^{\min} \le t_s \le t_s^{\max} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The resulting posterior, $\pi(t_s|t_m)$, is the product of the prior multiplied by the likelihood function and appropriately normalized.

$$\pi(t_s|t_m) = \frac{\pi(t_s)p(t_m|t_s)}{\int \pi(t_s)p(t_m|t_s)dt_s} = \begin{cases} \dfrac{p(t_m|t_s)}{\int_{t_s^{\min}}^{t_s^{\max}} p(t_m|t_s)dt_s} & \text{for} \quad t_s^{\min} \le t_s \le t_s^{\max} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The Bayesian estimator computes a single estimate, t_e, from the posterior by considering an objective cost function, $l(t_e, t_s)$, that quantifies the cost of erroneously estimating t_s as t_e. The Bayesian estimate minimizes the posterior expected loss, which is the integral of the cost function for each t_s, weighted by its posterior probability, $\pi(t_s|t_m)$.

$$t_e = f_l(t_m) = \arg\min_{t_e} \left[\int l(t_e, t_s) \pi(t_s|t_m) dt_s \right] \tag{5}$$

Notice that the optimal estimate, t_e, is a deterministic function of the measured sample $f_l(t_m)$ in which the subscript l reflects the particular cost/loss function.

For the MLE model, the estimator $f_{MLE}(t_m)$ is associated with the sample interval that maximizes the likelihood function, which can be derived from equation (1).

$$f_{MLE}(t_m) = \arg\max_{t_s} \lambda_{t_m}(t_s) = t_m \left[\frac{-1 + \sqrt{1 + 4w_m^2}}{2w_m^2} \right] \tag{6}$$

The MLE estimate is proportional to measurement. For a plausible range of values for w_m, the constant of proportionality would be less than 1, and thus the MLE estimator would systematically underestimate the sample. For example, for $0.1 < w_m < 0.3$, the constant of proportionality would vary between 0.99 and 0.92.

For the MAP rule, the cost function is $-\delta(t_e - t_s)$, where $\delta(.)$ denotes the Dirac delta function. The corresponding estimator function, $f_{MAP}(t_m)$, is specified by the mode of the posterior.

$$f_{MAP}(t_m) = \arg\max_{t_s} \pi(t_s|t_m) = \begin{cases} t_s^{\min} & \text{for} \quad t_s \le t_s^{\min} \\ f_{ML}(t_m) & \text{for} \quad t_s^{\min} \le t_s \le t_s^{\max} \\ t_s^{\max} & \text{for} \quad t_s \ge t_s^{\max} \end{cases} \tag{7}$$

For the BLS rule, the cost function is the squared error, $(t_e - t_s)^2$, and the estimator function, $f_{BLS}(t_m)$, corresponds to the mean of the posterior.

$$f_{BLS}(t_m) = \frac{\int_{t_s^{\min}}^{t_s^{\max}} t_s\, p(t_m|t_s)\, dt_s}{\int_{t_s^{\min}}^{t_s^{\max}} p(t_m|t_s)\, dt_s} \qquad (8)$$

The Bayesian observer model. The Bayesian estimator specifies a deterministic mapping from a measurement, t_m, to an estimate, t_e. However, our psychophysical data consists of pairs of sample interval, t_s, and production time, t_p. Accordingly, we augmented the estimator with a measurement stage and a production stage, which, together with the estimator, provides a complete characterization of the relationship between t_s and t_p. The model, however, relies on two intermediate variables, t_m and t_e, that are psychophysically unobservable (that is, hidden variables). To remove these variables from the description of the model, we took advantage of a trick common to Bayesian inference, which is to integrate out the hidden variables (that is, marginalization). Specifically, using the chain rule, we decomposed the joint conditional distribution of variables t_m, t_e and t_p to three intervening conditional probabilities:

$$p(t_p, t_e, t_m | t_s, w_m, w_p) = \\ p(t_p | t_e, t_m, t_s, w_m, w_p)\, p(t_e | t_m, t_s, w_m, w_p)\, p(t_m | t_s, w_m, w_p) \qquad (9)$$

We used the serial architecture of our model (**Fig. 3a**) to simplify the dependencies on the right hand side of equation (9). In the first term, because the conditional probability of t_p is fully specified by t_e and w_p (from equation (2)), we can safely omit the other conditional variables (t_m, t_s and w_m). In the second term, the only relevant conditional variable is t_m, as it specifies t_e deterministically. And for the third term, w_p has no bearing on t_m. Incorporating these simplifications, the joint conditional distribution can be rewritten as

$$p(t_p, t_e, t_m | t_s, w_m, w_p) = p(t_p | t_e, w_p)\, p(t_e | t_m)\, p(t_m | t_s, w_m) \qquad (10)$$

Moreover, because t_e is a deterministic function of t_m—that is, $t_e = f(t_m)$—the conditional probability $p(t_e | t_m)$ can be written as a Dirac delta function.

$$p(t_p, t_e, t_m | t_s, w_m, w_p) = p(t_p | t_e, w_p)\, \delta\big[t_e - f(t_m)\big]\, p(t_m | t_s, w_m) \qquad (11)$$

We can eliminate the dependence on the two hidden variables t_m and t_e by marginalization.

$$p(t_p | t_s, w_m, w_p) = \iint p(t_p, t_e, t_m | t_s, w_m, w_p)\, dt_m\, dt_e \\ = \iint p(t_p | t_e, w_p)\, \delta\big[t_e - f(t_m)\big]\, p(t_m | t_s, w_m)\, dt_m\, dt_e \qquad (12) \\ = \int p(t_p | f(t_m), w_p)\, p(t_m | t_s, w_m)\, dt_m$$

The integrand is the product of the conditional probability distributions associated with the measurement and production stages. By substituting these distributions from equations (1) and (2), and $f(t_m)$ from equations (6), (7) or (8)

(depending on the estimator of interest), equation (12) provides the conditional probability of t_p for a given t_s as a function of the model parameters w_m and w_p.

The Bayesian observer-actor model. The observer model described above obtains an estimate that minimizes a cost built around the estimate and the actual time interval. It was formulated to minimize the expected loss associated with erroneous estimates, rather than production times. A more elaborate Bayesian observer-actor model would seek to minimize expected loss with respect to the ensuing production times (and not the intervening estimates). This elaboration demands two considerations. First, the uncertainty associated with both the measurement and the production phases must be taken in to account. As such, the relevant probability distribution would be the joint posterior of the sample interval and production time conditioned on the measurement, $\pi(t_p, t_s | t_m)$. Second, the definition of the cost function should concern the sample interval and production time; that is, $l(t_p, t_s)$. The appropriate posterior expected loss could then be minimized as

$$t_e = f_l(t_m) = \arg\min_{t_e}\left[\iint l(t_p, t_s)\pi\big(t_p, t_s | t_m\big)\, dt_s\, dt_p\right] \qquad (13)$$

The delta and least-squares cost functions in this optimization problem do not correspond to the mode and mean of the joint posterior and derivation of the optimal solution is more involved and beyond the scope of our study. Nonetheless, we note that the corresponding estimators for the Bayesian observer-actor are qualitatively similar to those we derived for the MAP and BLS mapping rules in our simplified Bayesian observer model.

Fitting the model to the data. We assumed that t_p values associated with any t_s were independent across trials and thus expressed the joint conditional probability of individual t_p values across all N trials and across the three prior conditions by the product of their individual conditional probabilities.

$$p(t_p^1, t_p^2, t_p^3, \ldots, t_p^N | t_s, w_m, w_p) = \prod_{i=1}^{N} p(t_p^i | t_s, w_m, w_p) \qquad (14)$$

The products change to sums by taking the logarithm of both sides.

$$\log p(t_p^1, t_p^2, t_p^3, \ldots, t_p^N | t_s, w_m, w_p) = \sum_{i=1}^{N} \log p(t_p^i | t_s, w_m, w_p) \qquad (15)$$

Each term in the sum was derived from equation (12), after substituting $f(t_m)$ with the appropriate estimator function (equations (6), (7) or (8)).

We used this equation to maximize the likelihood of model parameters w_m and w_p across all t_s and t_p values measured psychophysically. The maximization was done using the fminsearch command in MATLAB (Mathworks), which incorporates the Nelder-Mead downhill simplex optimization method. Integrals of equations (8) and (12) are not analytically solvable and were thus approximated numerically using the trapezoidal rule. We evaluated the success of the fitting exercise by repeating the search with different initial values; the likelihood function near the fitted parameters was highly concave and the fitting procedure was stable with respect to initial values.