Dopamine firing plays a dual role in coding reward prediction errors and signaling motivation in a working memory task

Stefania Sarno[a,b,c,d,1] (ORCID), Manuel Beirán[a,e,2], Joan Falcó-Roget[a,b,2], Gabriel Diaz-deLeon[f], Román Rossi-Pool[f,g] (ORCID), Ranulfo Romo[f,g,h,1], and Néstor Parga[a,b,1]

[a]Departamento de Física Teórica, Universidad Autónoma de Madrid, 28049 Madrid, Spain; [b]Centro de Investigación Avanzada en Física Fundamental, Universidad Autónoma de Madrid, 28049 Madrid, Spain; [c]Aix-Marseille Université, CNRS, INSERM, Institut de Neurobiologie de la Méditerranée (INMED, UMR S901), Turing Centre for Living Systems, 13009 Marseille, France; [d]Aix-Marseille Université, CNRS, Institut de Recherche sur les Phénomènes Hors Équilibres (IRPHE, UMR 7342), 13013 Marseille, France; [e]Laboratoire de Neurosciences Cognitives Computationnelles, École Normale Supérieure – Paris Sciences et Lettres University, INSERM U960, 75005 Paris, France; [f]Instituto de Fisiología Celular, Departamento de Neurociencia Cognitiva, Universidad Nacional Autónoma de México, 04510 México City, México; [g]Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, 04510 Ciudad de México, México; and [h]El Colegio Nacional, 06020 México DF, México

Little is known about how dopamine (DA) neuron firing rates behave in cognitively demanding decision-making tasks. Here, we investigated midbrain DA activity in monkeys performing a discrimination task in which the animal had to use working memory (WM) to report which of two sequentially applied vibrotactile stimuli had the higher frequency. We found that perception was altered by an internal bias, likely generated by deterioration of the representation of the first frequency during the WM period. This bias greatly controlled the DA phasic response during the two stimulation periods, confirming that DA reward prediction errors reflected stimulus perception. In contrast, tonic dopamine activity during WM was not affected by the bias and did not encode the stored frequency. More interestingly, both delay-period activity and phasic responses before the second stimulus negatively correlated with reaction times of the animals after the trial start cue and thus represented motivated behavior on a trial-by-trial basis. During WM, this motivation signal underwent a ramp-like increase. At the same time, motivation positively correlated with accuracy, especially in difficult trials, probably by decreasing the effect of the bias. Overall, our results indicate that DA activity, in addition to encoding reward prediction errors, could at the same time be involved in motivation and WM. In particular, the ramping activity during the delay period suggests a possible DA role in stabilizing sustained cortical activity, hypothetically by increasing the gain communicated to prefrontal neurons in a motivation-dependent way.

dopamine activity | internal bias | ramping DA | working memory | motivation

Dopamine (DA) neurons in midbrain areas play a crucial role in reward processing and learning (1). Multiple studies have demonstrated that, in simple conditioning paradigms, phasic bursts in the DA firing encode the discrepancy between the expected and actual reward—the reward prediction error (RPE) (1, 2). Recent work has shown that stimulation uncertainty affects the phasic responses of midbrain DA neurons to relevant cues during perceptual decision-making tasks. It was observed that those responses were influenced by cortical information processing such as stimulus detection (3) and evidence accumulation (4). Importantly, under uncertain conditions, phasic DA responses contain information about the animal's understanding of the environment [its "belief state" (5–10)] and depend on the level of uncertainty that the animal has on its choice (3, 5, 6). While the influence of uncertainty in DA response had been, for long, overlooked, it was shown that the responses do not contradict the prevailing model for DA response, the RPE hypothesis (5, 6), but represent extensions of this model.

Although these studies on uncertainty represent substantial progress toward understanding how DA neurons respond, little is known about phasic and tonic (fast and slow fluctuations) DA signaling in decision-making tasks more cognitively demanding than simple stimulus detection or evidence accumulation. It is known that DA plays other roles beyond representing RPE (11). For example, in tasks involving working memory (WM), manipulations of the DA system crucially affect the behavioral performance and alter the activity of cortical areas implicated in WM such as the prefrontal cortex (12). Also, DA firing rate can reflect the motivation to work for reward (13). However, how these purported functions could affect decision-making tasks was not investigated.

The vibrotactile frequency discrimination task (14) facilitates deeper investigation into these issues. In this task, monkeys have to discriminate between two randomly selected vibrotactile stimulus frequencies, which are individually paired for each

Significance

Recent studies have confirmed the role of dopamine firing in reward prediction error, even under perceptual uncertainty. However, little is known about dopamine behavior during the use of working memory or its role in motivation to work for reward. Here, we investigated these issues in a discrimination task. Fast dopamine responses reflected a perceptual bias while remaining consistent with the reward prediction error hypothesis. When the bias increased task difficulty, motivation positively correlated with both performance and dopamine activity. In addition, dopamine slowly ramped up in a motivation-dependent way during the working memory period. Characterizing dopamine neurons' activity during tasks in which motivation influences behavior could importantly advance our knowledge of dopamine roles in effortful control.

NEUROSCIENCE

trial and separated by 3 s (the delay; Fig. 1A). Importantly, the monkeys have to maintain a trace of the first stimulus (f_1) during the delay period for comparison against the second stimulus (f_2) to reach a decision: f_1 is greater than f_2 ($f_1 > f_2$), or vice versa ($f_1 < f_2$). The presence of a delay between the two stimuli has two important implications that make this task challenging: First, the reliance on WM, and second, the influence of prior knowledge that makes the representation of the first stimulus shift toward the center of its distribution, a phenomenon called contraction bias (15).

We focused our analysis on three main points. The first is whether and how the contraction bias affects the DA activity. This is a question of interest because the contraction bias influences the trial difficulty in many instances of perceptual discrimination in both humans and rodents (15–17). The bias could thus influence reward expectations and, consequently, the phasic DA responses (given that the trial difficulty determines the reward expectation, which, in turn, shapes the DA response). Furthermore, since the bias is likely to originate from WM (15, 18), it is reasonable to ask whether and how it affects the DA delay-period activity.

The second is whether and how the DA activity during the WM period was temporally modulated and/or sensitive to the actual remembered stimulus frequency (f_1). This is a question of interest because it is known that DA neurons may be associated with frontal area responses during WM and, during the same task, frontal cortex neurons were tuned to the identity of f_1 during the WM delay in a time-dependent manner (14, 19, 20). Also, although it is believed that delay-period DA has a stabilizing role in prefrontal cortex sustained activity, there is little information about the DA firing activity during that period and the current view is that it remains at its baseline (21, 22).

Finally, the third is whether and how animal motivation relates to behavioral performance and to the DA time course. This is a question of interest because DA is believed to play a major role in motivated behavior (23) that, in turn, is expected to affect behavioral performance. In our task, the motivational level of the animal can be indirectly measured at the beginning of each trial by considering the time it takes to react to the start cue signal (reaction time, RT; Fig. 1A). Following this logic, Satoh et al. (13) found that the phasic DA activity, besides conveying an RPE signal, strongly correlated with motivation (see also ref. 23). On the other hand, prior claims associated tonic DA activity with conveying motivation (24, 25). Despite all these observations, the relationship between motivation, performance, and DA phasic/tonic signals has been poorly investigated in decision-making tasks.

We found that the DA activity simultaneously coded reward prediction and motivation. Phasic responses to the stimuli reflected the contraction bias in a way consistent with the RPE hypothesis. The activity during the WM period, although insensitive to the actual remembered stimulus, remained constantly above baseline and positively correlated with motivation (as measured by the RT). Similar positive correlations with motivation were found in the phasic responses to the starting cue and the first stimulus. At a behavioral level, higher motivation was indicative of improved performance, an effect that in our model-based analysis derived from a more accurate memory of the first stimulus.

Overall, our study shows that, in addition to encoding bias-modulated RPEs, the signal carried by the firing of DA neurons relates to motivation and might potentially play an important role in WM.

Results

Contraction Bias and Motivation Shape Behavioral Performance.
Monkeys were trained to discriminate between two vibrotactile stimuli with frequencies within the flutter range (Fig. 1A). Each stimulus pair (f_1, f_2; class) consisted of a large, unambiguous difference (8 Hz in most classes; Fig. 1B). In Fig. 1B, we show that performance for each class was only partially predictable from the absolute value $\Delta f = |f_1 - f_2|$. The contraction bias could explain this disparity, since the decision accuracy depended on the specific stimuli that were paired. This internal process shifts the perceived frequency of the base stimulus f_1 toward the center of the stimulus frequency range ($<f_1> = 21.5$ Hz; Fig. 1C). As such, if f_1 has a low stimulus value (i.e., 10 Hz), it will be perceived as larger, while a large frequency for f_1 is perceived as smaller. Therefore, correct evaluations for classes with $f_1 < f_2$ are increasingly obfuscated by the bias's effect as classes are visited from right to left (upper diagonal, Fig. 1 B and C), while classes with $f_1 > f_2$ are increasingly facilitated by the bias (lower diagonal, Fig. 1 B and C). To preserve this bias structure, we labeled classes with beneficial bias effects as closer to the extreme ends, and classes with hindrance from the bias as the center classes (Fig. 1C). In this class organization, decision accuracy was highest for the classes near the two ends (bias benefit) and, when plotted as a function of class number, the resulting curve was U-shaped (Fig. 1D).

To determine whether the motivation of the animals affected their behavioral performance, we analyzed the influence of RT on accuracy. First, we classified trials according to RT into two groups (short and long; *SI Appendix*) and observed that, in both conditions, the distribution of trials as a function of class number was not significantly different from the uniform distribution (Fig. 1E; chi-squared test, $P = 0.46$). This result confirmed that, at the beginning of the trial, the animal had no clue about the upcoming class and thus that the RT can be considered as a behavioral measure of motivation independent of the difficulty.

Since motivation can affect perception (26), we next considered how the RT was related to accuracy. We observed that accuracy for long RT was lower than for short RT (Fisher's exact test, $P = 0.036$). As a function of class number (Fig. 1 F, Left), in both short- and long-RT trials, the accuracy decreased for classes disfavored by the bias (classes 4 to 9; Fig. 1C, purple). Furthermore, in those same classes, the performance was significantly better for short-RT trials (Fig. 1 F, Right; Fisher's exact test, $P = 0.003$); on the contrary, differences in accuracy did not reach a significant level (Fisher's exact test, $P = 0.88$) in classes that were favored by the bias (classes 1 to 3 and 10 to 12; Fig. 1C, pink). Finally, we compared the RT in correct and error trials (Fig. 1G) and observed that the RT was significantly shorter in correct trials (t test, $P = 0.009$).

All the above results indicated that the motivation had a clear and significant effect on the performance of the animal: It correlated with the trial outcome and significantly improved the accuracy in difficult trials (those corresponding to classes disfavored by the bias).

DA Responses Reflect the Contraction Bias.
We recorded single-unit activity from 22 putative midbrain DA neurons while trained monkeys performed the discrimination task. They were identified based on previous electrophysiological criteria (27): regular and low tonic firing rates (mean ± SD: 4.7 ± 1.4 spikes per second), a long extracellular spike potential (2.4 ± 0.4 ms), positive activation to reward delivery in correct (rewarded) trials, and with a pause in error (unrewarded) trials (28).

We found that the majority of neurons (16 out of 22, 72% of the population) showed a significant, positive modulation to at least one of the two stimuli (*SI Appendix*). We therefore analyzed how the DA neurons responded during the stimulation period. It is known that, in frontal lobe neurons, the animal's decision is observable during the comparison period [i.e., during the presentation of f_2 (20)]. We hypothesized that if DA neurons receive information from the frontal lobe, their responses to f_2 should reflect the contraction bias's effect. As a result, neurons showed significant dependence on stimulus classes (one-way ANOVA, $P = 0.0013$; Fig. 2A). The U-shaped
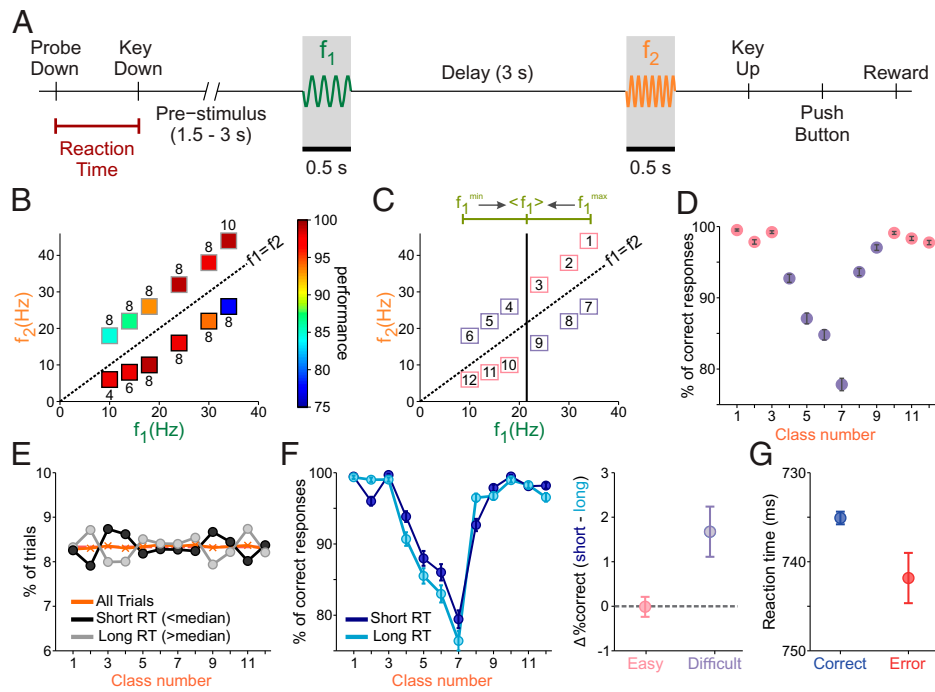
Fig. 1. Discrimination task and contraction bias. (*A*) Schematic of the task design. The trial begins with the lowering of the mechanical stimulation probe (probe down). The monkey reacted by placing its free hand on an immovable key (key down). The time interval between these two events is known as the reaction time (RT, underlined in red). After a variable period (1.5 to 3 s), the probe oscillated for 0.5 s at the base (f_1) frequency. A 3-s delay after the first stimulus constitutes the working memory (WM) period. The delay is followed by the presentation of the second stimulus (f_2, also the "comparison period"). The offset of f_2 signals the monkey to report his decision by pressing one of the two push buttons (PBs) to indicate whether the comparison frequency was higher or lower than the base frequency. After that, the animal was rewarded for correct discriminations, or received a few seconds of delay for incorrect discriminations. (*B*) Stimulus set composed of frequency pairs (f_1, f_2) used in the task. Frequency values for f_1 are on the *x* axis, while values for f_2 are on the *y* axis. The diagonal dashed line represents stimulus equivalence ($f_1 = f_2$). The box colors indicate the percentage of correct trials, while the numbers above or below each box indicate the absolute magnitude difference (Δf) for each class. Red is for the highest correct percentages, and blue is for the lowest. (*C*) The upper line (green) depicts the contraction bias effect, where the base stimulus is perceived as closer to the mean value of the first frequency ($<f_1>$). The solid, vertical black line marks this center value. Numbers label classes, with the diagonal line marking stimulus equivalence (black, dashed). The upper diagonal ($f_1 < f_2$) has bias benefit (pink squares) above the range center, and bias obstruction (purple squares) is below the range center. The lower diagonal ($f_1 > f_2$) has bias benefit below the range center (pink), and bias obstruction above the range center (purple). (*D*) The numbered classes (*x* axis) from *C* by performance percentage (% of correct responses), representing the color values used in *B*. Pink and purple circles represent classes with bias benefit and bias obstruction, respectively. Error bars (in black) represent SDs and were obtained by computing the performance 1,000 times, resampling with replacement from the original data. (*E*) Percentage of trials belonging to each class number for short-RT trials (RT < median) and long-RT trials (RT > median) (black and gray circles and lines, respectively). The orange line and crosses represent the percentage of trials in each class when all trials are considered. (*F*, *Left*) Percentage of correct trials, as a function of class number, for short and long RT (dark and light blue circles and lines, respectively). Error bars are computed as in *D*. (*F*, *Right*) Differences in the percentage of correct trials between long- and short-RT trials in easy and difficult classes (respectively, pink and purple circles and bars). (*G*) Mean RT in correct (blue) and error (red) trials. Error bars represent the SE.
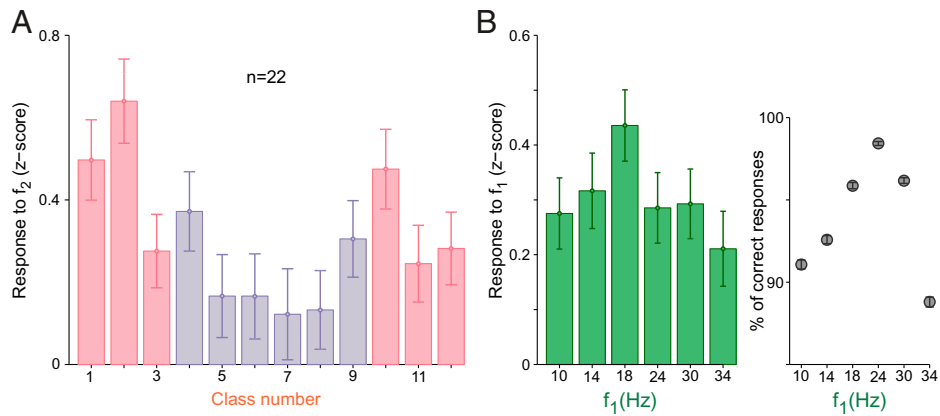
modulation in the response distributions can be explained by a pair of arguments: First, the bias induced a difficulty that affected belief states. Second, as an RPE signal, the DA response to f_2 coded the change in reward expectation produced by the application of the comparison frequency, which is higher for classes facilitated by the bias (Fig. 2*A*). A reinforcement learning model based on belief states supports this interpretation (*SI Appendix*, Fig. S1*A*). Similar effects of belief states on responses to reward-predicting stimuli were observed in other experiments (4–6). However, in those studies, the trial difficulty was controlled by experimental design while it was determined by a perceptual bias in our task.

Next, we wondered whether the same bias effect was observable during f_1. By design, each f_1 value is the same for two classes ($f_1 > f_2$ or $f_1 < f_2$), where one class is favored by the bias and the other is disfavored (Fig. 1*D*). However, as the animal had no clue about the upcoming class during the presentation of the first stimulus, it is reasonable to study the monkeys' performance and DA responses based on the f_1 value only. We observed that the decision accuracy was lowest for the extremum f_1 intensities (Fig. 2 *B*, *Right*). As we would expect, the phasic DA response to f_1 showed a similar modulation (Fig. 2*B*), although this effect did

not reach statistical significance (one-way ANOVA, $P = 0.279$). Nonetheless, DA responses for the middle values of our stimulus range (18 and 24 Hz) were significantly greater than those produced after the application of extreme values of the stimulus range (i.e., for 10 and 34 Hz; two-sample one-tailed *t* test, $P = 0.039$).

These results show that DA responses did not encode physical properties of the stimuli. Instead, they coded reward expectations as induced by the contraction bias.

A Bayesian Model Replicates the Contraction Bias and Animal Accuracy. We reasoned that the contraction bias could influence the animal's difficulty with a class, similar to how the stimulus physical features do (see, e.g., ref. 29). To analyze this issue, we used a Bayesian approach in which observation probabilities were combined with prior knowledge to obtain a posterior probability or belief about the state of the stimuli (16, 18, 30, 31) (Fig. 3*A*). The prior probabilities of f_1 and f_2 were assumed to be discrete and uniform, and we supposed that animals only perceived noisy representations of the two stimuli (Fig. 3*A*). Observations of f_1 and f_2 were obtained from Gaussian distributions at the end of the delay period, with respective SDs σ_1 and

Sarno et al.
Dopamine firing plays a dual role in coding reward prediction errors and signaling motivation in a working memory task

PNAS | 3 of 10
https://doi.org/10.1073/pnas.2113311119

Fig. 2. (*A*) Normalized population responses to f_2, sorted by stimulus class (*SI Appendix, Methods*). Error bars are ±1 SEM. (*B, Left*) Normalized population response to f_1 plotted against f_1 frequencies. Error bars are ±1 SEM. (*B, Right*) Percentage of correct responses for each f_1 frequency.

σ_2 as noise parameters. Decisions were made using the belief about the state $f_1 > f_2$ and the maximum a posteriori (MAP) decision rule. The noise parameters were adjusted to optimize the similarity between the Bayesian model and the animal's performance (Fig. 3*B*; see *SI Appendix* for model and parameter fitting). The best fit yielded $\sigma_2 = 3.2$ Hz and $\sigma_1 = 5.50$ Hz (Fig. 3*B*), so the noise parameter for f_1 was greater than that for f_2; from this, we intuited that the representation of f_1 deteriorated over the WM period. When this happens, the Bayesian model produces a contraction bias because the inferred value of f_1 is noisier than that of f_2, and the inferred value of f_1 relies more heavily on its prior distribution (16).

We then used the fitted Bayesian behavioral model to simulate the discrimination task and to evaluate the difficulty for

each trial. Difficulty was defined as the decision uncertainty (i.e., the complement of confidence, as defined statistically) (29), given by the probability that the choice made by the model was not correct (Fig. 3*C*). Also, the difficulty associated with a class was defined as its average over many simulated trials in that class. In correct trials, class difficulty versus class number was an inverted U (Fig. 3*D*, blue line), increasing/ decreasing as the bias was less/more favorable to making a correct choice. Comparing the results between correct and error trials, the model had the opposite class difficulty pattern when it erred (Fig. 3*D*, red line).

To summarize: A simple encoding model replicated the contraction bias. According to the model the bias appears because the memory of the first stimulus deteriorates through the delay
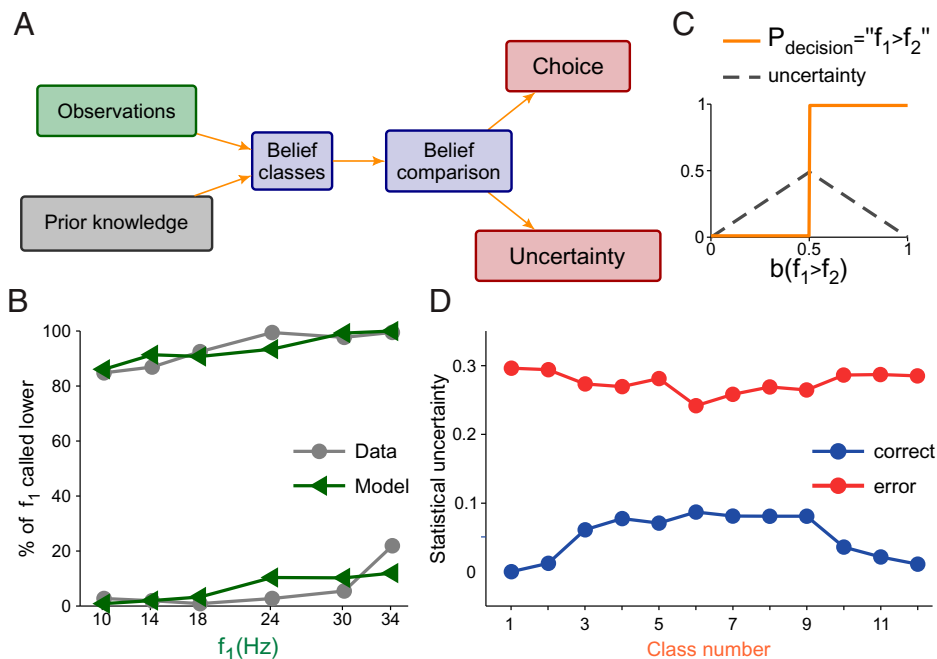
Fig. 3. Bayesian behavioral model and the relationship between the contraction bias and difficulty. (*A*) Schematic of the Bayesian behavioral model: Prior knowledge and observations are combined to obtain two comparison beliefs, b(H) ($f_1 > f_2$; *Higher*) and b(L) ($f_1 < f_2$; *Lower*). These beliefs are compared by magnitude in order to reach a decision. The decision can be decomposed into two components: choice and uncertainty. See *SI Appendix, Methods* for a more accurate description of the model. (*B*) Model fit of the percentage of times in which the base frequency (f_1) is called smaller. The behavioral data are in gray (circles, lines), and model predictions are in green (triangles, lines). Line pairs for the 12 classes divided by the diagonal in Fig. 1*D* ($f_1 < f_2$, *Upper* pair; $f_1 > f_2$, *Lower* pair). Model best-fit parameter values were $\sigma_1 = 5.5$ Hz and $\sigma_2 = 3.2$ Hz. (*C*) Choices are made using the MAP criterion (orange line). Uncertainty is defined as 1 minus the maximum of the two beliefs b(H) and b(L) (gray line). (*D*) Bayesian model uncertainty predictions: Correct trials are in blue (circles, lines), and wrong trials are in red (circles, lines).

period. This mechanism eventually degrades the performance in a nontrivial way (Figs. 1D and 3B). An important consequence of these results is that, in our task, class difficulty is mainly modulated by an internally generated process, and not by the physical properties of the stimuli.

Motivation Improved Perceptual Precision. We have previously noticed that the performance was better in short-RT trials, especially in classes negatively affected by the bias (Fig. 1F). We thus explored how this performance enhancement arose in the Bayesian framework by fitting the model parameters to both trial groups independently (*SI Appendix*, Fig. S2). We observed that the noise parameter σ_1 (emulating the uncertainty in the representation of the memorized frequency f_1) was smaller in the short-RT condition (*t* test, $P < 0.001$). A smaller σ_1 decreases the contribution of the prior distribution to the posterior of f_1, resulting in a weaker contraction bias.

Thus, in short-RT trials, the performance improved in classes disfavored by the bias. Instead, in classes favored by the bias, the reduction of the uncertainty parameter had almost no effect, because the discrimination is already very easy.

Difficulty Shapes DA Responses during Decision Formation. To study the influence of class difficulty on DA activity, we considered correct trials and subdivided the stimulus set (12 classes) into two groups: a low-difficulty group (classes 1 to 3 and 10 to 12; Fig. 1C, pink) and a high-difficulty group (classes 4 to 9; Fig. 1C, purple). During presentation of f_1, the normalized activity (*z* score) showed a tendency to be higher in the low-difficulty group (Fig. 4 A, Left); however, this tendency was not found to be significant ($P = 0.17$, two-tailed *t* test; *SI Appendix*). Importantly, the difference became significant during the presentation of f_2 (Fig. 4 A, Center; two-tailed *t* test, $P < 0.001$; *SI Appendix*). We obtained a response latency of 245 ms for the time it took the two difficulty groups to significantly diverge in responses after f_2 onset (area under the receiver operating characteristic curve [AUROC], $P < 0.05$; sliding 250-ms window, 10-ms steps; *SI Appendix*). Dependence on class difficulty during decision formation is similar to the dependence on choice confidence as reported in previous work (5, 6). Interestingly, the activity after delivery of reward was invariant to the difficulty level, having the same peak response for both class groups (Fig. 4 A, Right).

The Bayesian model predicted higher uncertainty for error trials (Fig. 3D), so we studied the phasic activity, sorting trials by decision outcome (correct vs. incorrect). Responses to f_1 did not vary based on outcome (Fig. 4 B, Left), but the responses did vary during the comparison stimulus. The temporal profiles for each outcome, analogous to the difficulty groups, separated with an approximate latency value of 205 ms (Fig. 4 B, Center). This time lag in DA signals is comparable to those found in the secondary somatosensory cortex (32, 33) and frontal areas during the same task (20, 34). Similarities in latency between the frontal lobe and DA neurons were also found during a tactile detection task (35).

At the point of reward delivery, the phasic DA activity increased in correct trials but decreased for error trials (Fig. 4 B, Right). This temporal pattern is consistent with that predicted by the RPE hypothesis. The latency of divergence of these two response profiles (AUROC, $P < 0.05$; *SI Appendix*) was shorter than those after f_2 onset (∼130 ms). Further, we asked whether a reinforcement learning model based on belief states could emulate the DA responses. The model reproduced the phasic responses observed in Fig. 4 well (*SI Appendix*, Fig. S1 B and C).
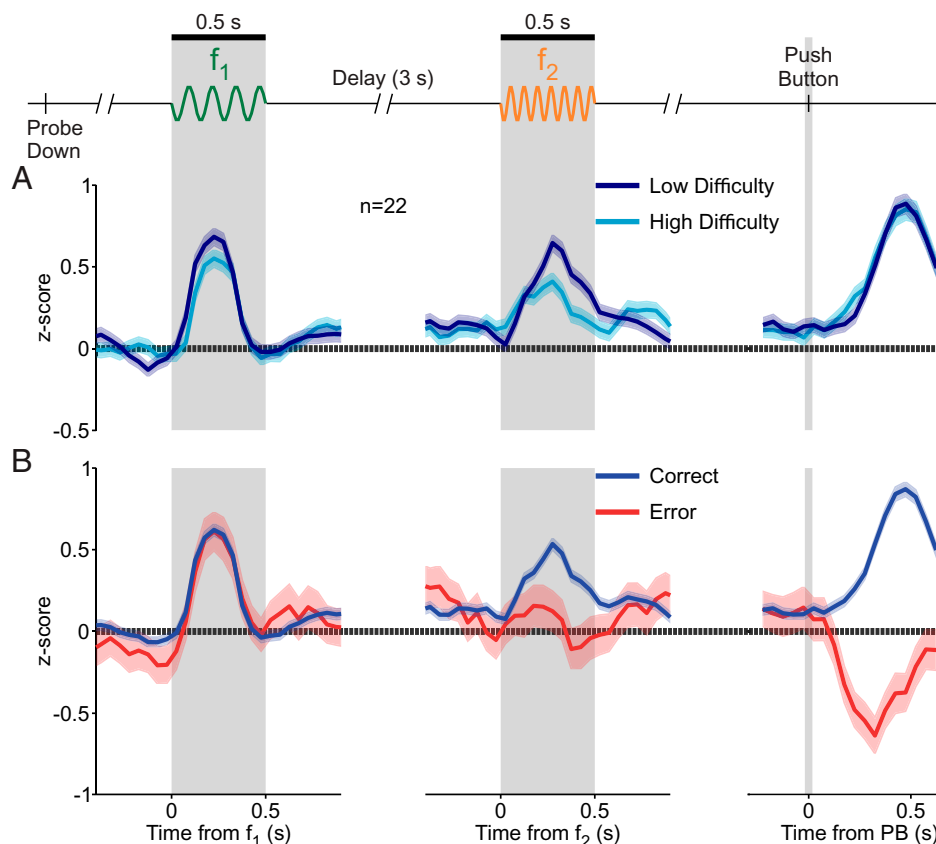
These results show that DA responses are compatible with RPEs. Furthermore, during the second stimulus, these responses reflect difficulty, as defined by the uncertainty in our Bayesian model (Fig. 3D).

DA Ramps Up during the Delay Period without Coding f_1. Because during the WM period, neurons in frontal areas exhibit a temporally modulated sustained activity, tuned parametrically to f_1 (14, 19, 20), we investigated how DA neurons behave during the same period. Is their activity modulated and tuned to f_1? Fig. 5 A, Left shows the normalized activity (*SI Appendix*) of an example neuron with significantly increased activity through the entire WM period. When we sorted the firing rates by f_1 values, we saw a clear ramping in activity across the delay period in all six curves (Fig. 5 A, Center). However, the temporal profiles are not modulated by the f_1 values. To further verify the absence of tuning, we averaged the entire delay activity for each f_1 value (Fig. 5 A, Right), and found that this neuron did not exhibit significant differences across f_1 (one-way ANOVA, $P = 0.76$). Fig. 5 B, Left shows another example neuron. Its activity exhibited absolutely no temporal modulation during WM, remaining closer to its baseline value (Fig. 5 B, Left). The absence of ramping persisted when we expanded the observations to differentiate between responses to each f_1 stimulus (Fig. 5 B, Center). Further, neither the temporal profile nor the averaged activity during WM (Fig. 5 B, Right) was modulated based on f_1 (one-way ANOVA, $P = 0.64$).

At the population level, we found that neurons responded to f_1 by increasing their firing rate (one-tailed Wilcoxon signed-rank test, $P < 0.001$; Fig. 5C) in a way that was not tuned to the value of the frequency. Furthermore, dividing the WM period into three segments (Fig. 5C; black horizontal bars indicate in the delay period), we observed positively and significantly modulated activity in all of them when compared with a segment before the presentation of f_1 (Fig. 5C, pink bar; one-tailed Wilcoxon signed-rank test, $P = 0.012$, $P < 0.001$, and $P < 0.01$, respectively). The average population activity exhibited a ramping increase during the WM period (Fig. 5D). Note, however, that, during the last portion of the delay period, the DA signal decreased, resembling a "bump," possibly for the effects of temporal uncertainty (36).

Due to the fact that frontal neurons can be monotonically tuned to the identity of f_1 during WM, we sought to confirm that this coding did not exist within the DA population in any time interval. A linear regression in five discrete intervals (pink and black bars in Fig. 5 C, Top) showed no significant linear trend (Fig. 5E). Moreover, we extended our analysis to 1 s before f_1 onset through the end of the delay period (a sliding 250-ms window with 10-ms steps). In these 4.5 s, we performed linear and sigmoidal regressions in each window and found no significant results for either linear or sigmoidal regression, in any window (slope different from zero, $P < 0.01$, and fit with $Q > 0.05$; *SI Appendix*). Notably, the absence of monotonic dependence was held even during the last portion of the delay interval. This contrasts with the f_1 information recovery observed in frontal neurons (14), a result also found in other tactile tasks (20, 34, 37). Pure temporal ramping signals during WM were also identified in frontal areas in this and other tactile tasks (19, 38).

To exclude the existence of more general dependencies on f_1 during WM (e.g., similar to those in Fig. 2), we searched for temporal windows with significant ANOVA results ($P < 0.05$; *SI Appendix*). We focused on the same 4.5 s and found that in several moments during the presentation of f_1 and throughout the delay period the activity depended on f_1; however, this dependence was intermittent without exhibiting a clear or consistent pattern (*SI Appendix*, Fig. S3 A and B). This led us to seek temporal consistency, so we divided the entire time range into nine discrete subintervals of 500 ms and calculated the fraction of significant sliding windows in each (*SI Appendix*). We found a consistent dependence on f_1 only during the presentation of the stimulus, where the *z* scores depended on

Sarno et al.
Dopamine firing plays a dual role in coding reward prediction errors and signaling motivation in a working memory task

PNAS | 5 of 10
https://doi.org/10.1073/pnas.2113311119

Fig. 4. Phasic activation of DA neurons to the base stimulus, comparison stimulus, and reward delivery. Normalized DA population responses to different periods of the task (shadows ±1 SEM). (*Left*) Responses to f_1 (gray shaded region), aligned to f_1 onset. (*Center*) Responses to f_2 (gray shaded region), aligned to f_2 onset. (*Right*) Responses to reward delivery, after the decision report (slender gray region), aligned to PB. (*A*) Responses for difficulty (uncertainty) groups: low (navy blue) and high (light blue) difficulty. (*B*) Responses for trial outcome groups: correct (blue) and error (red) trials.

f_1 with a profile similar to the one shown in Fig. 2 *B*, *Left* (*SI Appendix*, Fig. S3*B*).

After studying f_1 in isolation, we repeated a similar analysis to identify dependencies on the stimulus class. We implemented the ANOVA test (*SI Appendix*, Fig. S3*C*) and searched for those windows in which $P < 0.05$. We then calculated the temporal consistency in discrete subintervals of 500 ms, repeating the same procedure as for f_1. We found significant class-dependent activity only during f_2 stimulation, with a pattern similar to that in Fig. 2*A* (*SI Appendix*, Fig. S3*D*).

Overall, despite finding some variability for single-neuron analysis, the entire population activity showed a positive temporal modulation without carrying any information about the remembered frequency.

Phasic DA Activity Reflected Motivation on a Trial-by-Trial Basis Only before the Decision Formation. To understand the influence of motivation on a trial-by-trial basis, we first classified trials according to RT into two groups (short and long; *SI Appendix*) and analyzed the time courses of DA activity using this classification (Fig. 6*A*). We noticed that the phasic responses to probe down (PD) and to f_1 were higher for short-RT trials (Fig. 6 *A*, *Left* and *Middle*). In contrast, both groups reached a common peak response to the second stimulus (Fig. 6 *A*, *Middle*) and to the reward delivery (Fig. 6 *A*, *Right*).

This analysis suggests that, prior to the presentation of f_2 and to the decision formation, the phasic DA responses to the PD and f_1 coded motivation (Fig. 6*A*). To confirm this result, we computed the correlation coefficient between the RTs and the phasic DA responses. We obtained a significant negative correlation between the DA responses to the PD and the RTs ($r = -0.07$, $P < 0.01$ with permutation test; Fig. 6 *B*, *Left* and *SI Appendix*). A similar result was found when we analyzed the correlations between the RTs and the DA activity after the first stimulus ($r = -0.05$, $P = 0.021$ with permutation test; Fig. 6 *B*, *Right*). In contrast, the correlations were not significant when we considered the responses to the second stimulus and reward delivery. Thus, the activity of DA neurons correlated with motivation at the start cue signal as well as throughout the first stimulation and WM periods of the task. Instead, during and after decision formation, the signal showed no interaction with motivation and reflected only the RPE that one would expect after each choice and decision outcome.

Tonic DA Activity Increasingly Correlated with Motivation throughout the WM Period. We finally analyzed how the DA activity varied with the RT during the WM period. We found that for short-RT trials, the tonic DA signal (z score) was consistently greater than for long-RT trials. The red bar in Fig. 6*A* indicates periods of significant differences between these two signals (AUROC, $P < 0.05$). When we averaged the activity for the RT groups across the whole WM period, the activity of short-RT trials was significantly higher ([mean ± SEM] short: $0.25 ± 0.03$; [mean ± SEM] long: $0.11 ± 0.03$; two-sample one-tailed t test, $P < 0.001$).

Furthermore, during WM, the trial-to-trial correlations between the DA activity and the RT were significantly negative ($r = -0.10$, $P ≤ 0.001$ with permutation test; Fig. 6 *C*, *Left* and *SI Appendix*). More interestingly, we noticed that the DA activity in short- and long-RT trials became increasingly different as
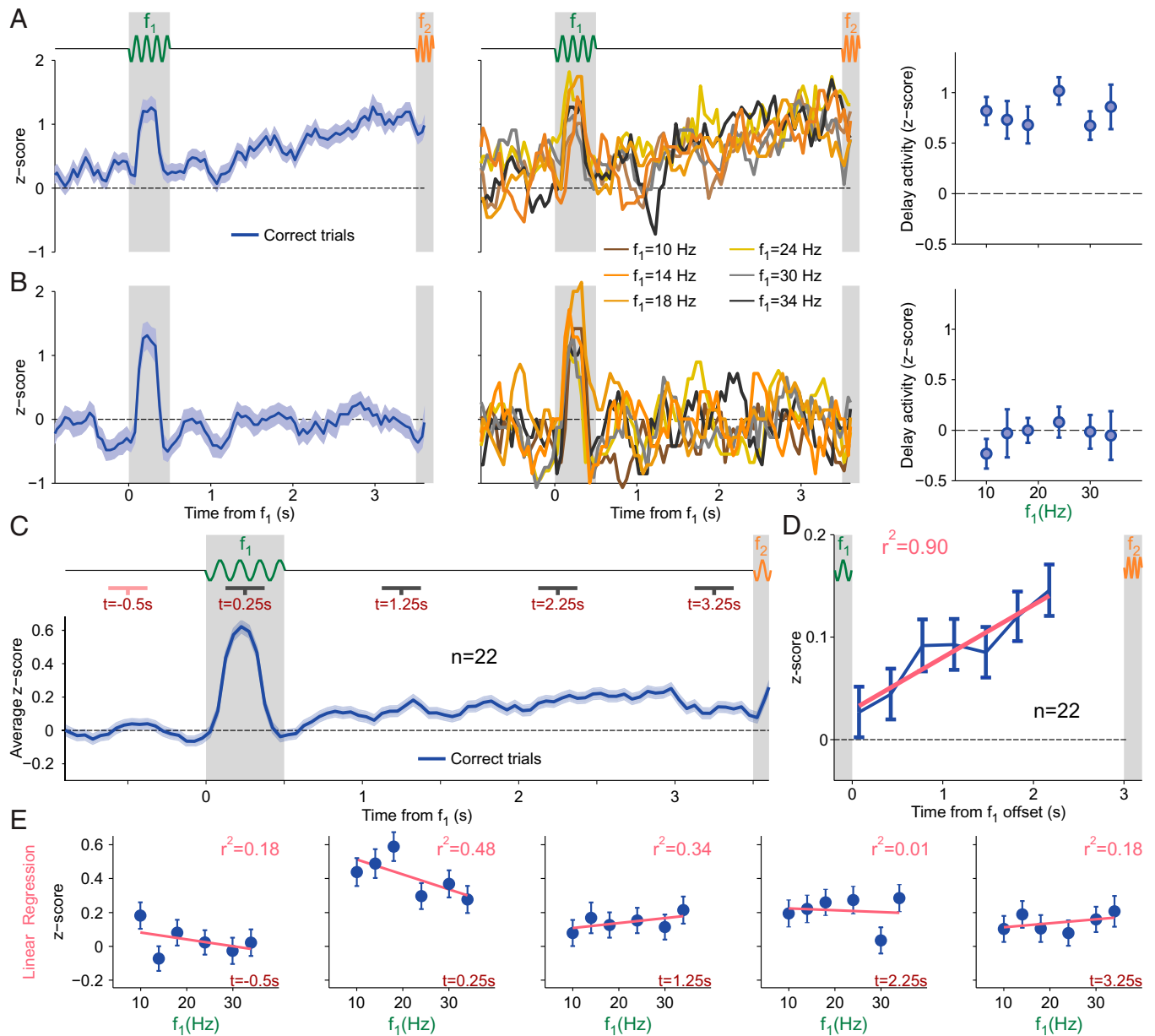
Fig. 5. DA activity during the delay period. (*A* and *B*) Normalized activity as a function of time for two example neurons; trials are aligned to f_1 onset. Activity is calculated in windows of 250 ms with steps of 50 ms (*SI Appendix, Methods*). Neither example shows coding activity during WM. (*A* and *B*, *Left*) Activity from all correct trials was averaged (blue line; shadows ±1 SEM). The thick gray shaded area is for f_1; the thin gray shaded region is for f_2. (*A* and *B*, *Center*) Same marked gray regions. Each curve indicates activity for one f_1 stimulus value. Curves with color closest to yellow are for intermediate values; the darkest curves are for the extreme values. (*A* and *B*, *Right*) Average activity during the WM (delay) period as a function of f_1 values (error bars ±1 SEM). (*A*) Exemplary neuron with temporal dynamics during WM. (*B*) Single neuron without temporal dynamic. (*C*) Normalized activity as a function of time between 1 s before f_1 onset through the end of WM (blue line; shadows ±1 SEM). Trials are aligned to f_1 onset. Gray regions mark the same periods as for *A* and *B*; activity is calculated in the same manner. Dark gray lines under task indicate analyzed periods after f_1 onset; the pink line under task indicates the basal period before f_1 onset. Pink periods compared against gray periods show significant modulations in each time segment (Wilcoxon signed-rank test, $P < 0.001$, $P = 0.012$, $P < 0.001$, and $P < 0.01$). (*D*) Mean population activity of DA neurons during the delay period, with correction for basal activity. Gray shadowed regions demonstrate stimulus presentation periods. Blue indicates the corrected firing rate in nonoverlapping 350-ms periods (±1 SEM). The pink line displays linear regression across seven discrete periods, where r^2 is the determination coefficient. We found a significant linear trend ($P = 0.001$). (*E*) Linear regressions (pink lines) calculated in each of the five indicated regions from *C*. Regressions over average activity for each f_1 stimulus (error bars ±1 SEM). The first graphic is for the basal period before f_1 onset, proceeding in increasing chronological order from left to right. Times are indicated in the lower right corner (dark red); r^2 values are the determination coefficients and show the accuracy of each linear fit to the data (upper right corner, light red). No linear regressions were found to have significant monotonic coding.

time elapsed during the delay period, showing a higher value in short-RT trials (Fig. 6*A*). To assess the significance of this effect, we divided the delay period into nonoverlapping bins and computed the correlations between the DA activity and the

RT in each of them (*SI Appendix*). We found that the correlation coefficient consistently decreased, becoming significantly negative during the last portion of the delay (Fig. 6 *C, Right* and *SI Appendix*). To summarize: The correlations between
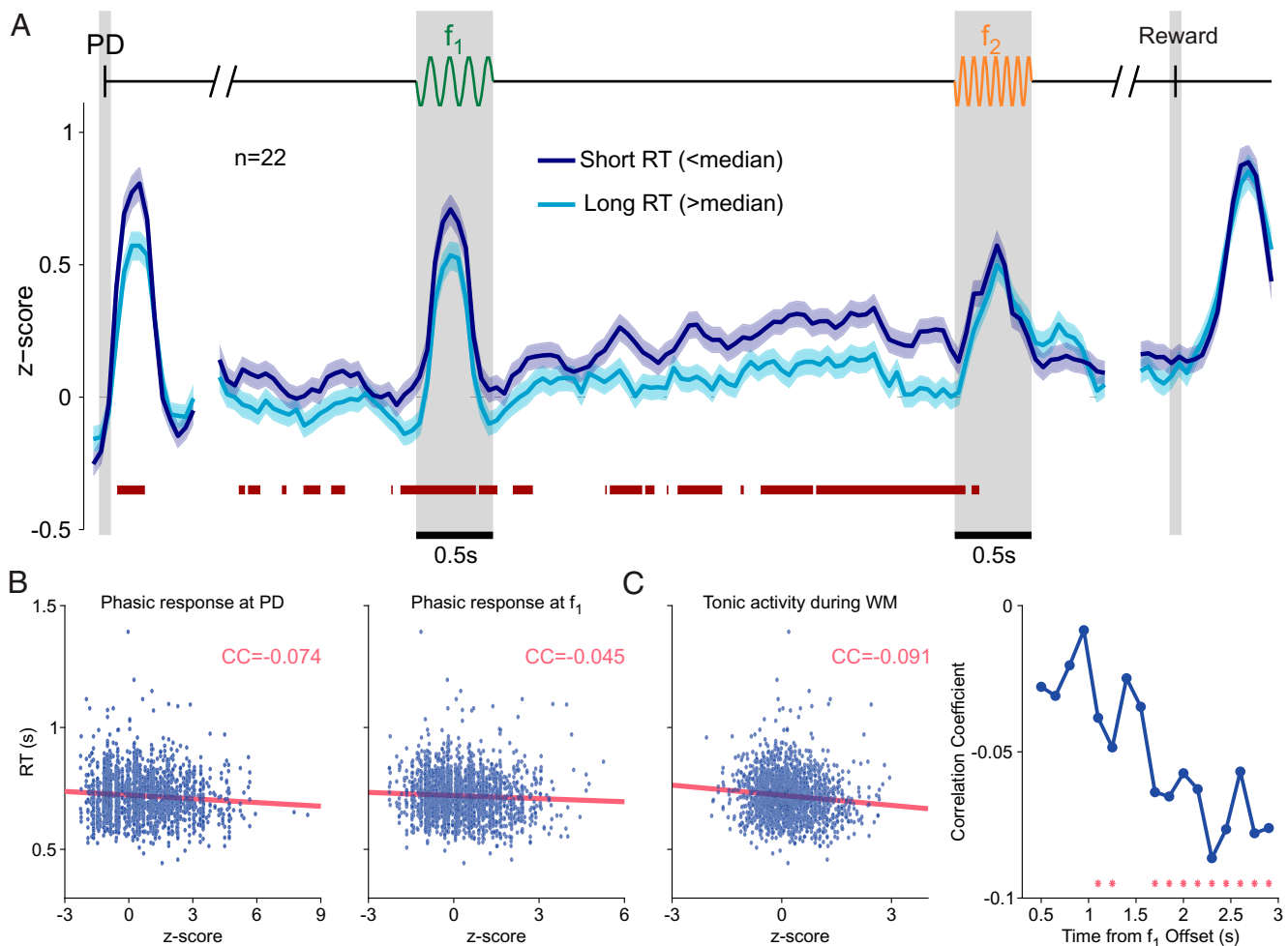
Fig. 6. DA activity in short- and long-RT trials. (*A*) Normalized population activity as a function of time for correct trials, divided by RT duration: short-RT trials (RT < median; dark blue line, shadows ±1 SEM) and long-RT trials (RT > median; light blue line, shadows ±1 SEM). The red bar below the curves marks significant differences between RT groups (AUROC, *P* < 0.05; *SI Appendix, Methods*). (*B*) Trial-to-trial correlation of the magnitude of the DA response to the PD (*Left*) and to f_1 (*Right*) with the RTs. (*C, Left*) Trial-to-trial correlation between the DA activity during WM and the RTs. (*C, Right*) Evolution of the correlation coefficient between the DA activity and the RTs through the WM period. Red stars mark windows with significant correlation coefficients. Note that correlations were consistently negative and became significant toward the end of the delay WM period.

motivation and DA activity were more pronounced at the end of the WM period, while DA firing simultaneously showed a positive temporal modulation.

Discussion

We sought to determine the activity of DA neurons recorded while monkeys performed a cognitively demanding discrimination task. We found that phasic DA responses before the comparison period coded RPE and motivation, although the DA activity during WM only coded motivation. We also observed that a subset of DA neurons exhibited a motivation-dependent ramping increase during the delay.

Contraction Bias and Motivation. The contraction bias affected the task difficulty. At the end of the delay period, observations of the first frequency have deteriorated, increasing the relevance of prior knowledge, and thus modulating performance (Fig. 1*D*) and DA responses (Fig. 2*A*). The effect of the bias seemed somewhat countered by motivation: For short-RT trials, the accuracy in difficult classes was improved, an effect likely due to a weakening of the contraction bias (Fig. 1*F*). Recently, Mikhael et al. (39) proposed that tonic DA controls how worthwhile it is to pay a cognitive cost that improves Bayesian

inference by increasing perceptual precision, ultimately leading to better performances.

Motivation and DA Activity. Motivation also influenced DA activity: The phasic responses to the start cue (the PD event) and the first vibrotactile stimulus depended on the RT, exhibiting a larger value for faster responses. Importantly, DA responses to those task events and the delay-period activity were negatively correlated with the RT, suggesting that DA drives motivated behavior on single trials, as it has been advocated in other quite different tasks (13, 23).

Ramping Activity. An interesting result was that DA neurons exhibited RT-dependent ramping activity during WM. This DA signal conveys motivational information, which could be playing relevant roles in cortical cognitive processes. Previous recordings in midbrain DA neurons during monkeys' performance of WM tasks did not observe sustained activity during the delay (21, 40, 41). This discrepancy can be attributed to differences in WM and attention requirements, as well as to task difficulty or delay duration. In the striatum, however, a ramping increase in the release of DA has been described (23, 24, 42, 43); this signal purportedly reflects an approaching decision report (24, 42), reward proximity

Sarno et al.
Dopamine firing plays a dual role in coding reward prediction errors and
signaling motivation in a working memory task

(43, 44), possibly aiding in action invigoration as well (45). Ramping in the spiking activity of DA neurons was observed in some experiments (44, 46) but not in others (42). Moreover, several theoretical explanations have been proposed for generating ramping signals (36, 47, 48). Ramping has been associated with the presence of external cues reducing uncertainty (36). Here we observed a ramping of DA activity in well-trained animals and in the absence of any external feedback able to reduce temporal or sensory uncertainty. The ramping DA signal appeared when an active maintenance of task-relevant information was required and, as such, we attribute its existence to the usage of WM.

Suggested Role of Delay-Period DA Activity. Although in our experiment we were not able to address the implications of the delay-period DA signal, we propose that this activity might be related to the cognitive effort needed to upload and maintain a percept in WM. Recent advances in the study of motivation and cognitive control (49) indicate possible directions to investigate this issue further in future work. Cognitive control is effortful (50, 51), involving a cost–benefit decision-making process. A task that involves WM requires an evaluative choice: either engaging in the task, or choosing a less demanding option (e.g., guessing), even if it leads to a smaller reward (45). An intriguing hypothesis relates motivation to cognitive control modulated by dopaminergic signals to prefrontal cortex and striatal neurons (49, 52, 53). In this scenario, DA is proposed as a persistent-activity stabilizer, increasing the gain in target neurons, and in turn producing an enhanced signal-to-noise ratio and promoting WM stability (54).

Conclusions. In summary, our results underline the essential role of midbrain DA neurons in learning, motivation, and WM under perceptual uncertainty. In our frequency discrimination task, phasic DA responses were affected by the contraction bias and motivation, while delay-period activity could instead be compatible with motivational behavior. The dependence of DA activity on the contraction bias suggests that internally generated biases can play a role in learning. Two-interval forced-choice tasks, such as discriminating between two temporal intervals, exhibit the contraction bias; however, DA was not investigated in these cases. Future study within these task paradigms can further our understanding of DA activity in motivation, cognitive control, and WM. Altogether, our results point to an intricate relationship between DA, perception, and WM, as they are all modulated by the animal's intrinsic motivation.

Materials and Methods

Monkeys were trained to report which of two vibrotactile stimuli was greater in frequency (*SI Appendix*). Neuronal recordings were obtained in cortical areas while the monkeys performed the vibrotactile frequency discrimination task. Animals were handled in accordance with standards of the NIH and Society for Neuroscience. All protocols were approved by the Institutional Animal Care and Use Committee of the Instituto de Fisiología Celular, Universidad Nacional Autónoma de México.

1. W. Schultz, P. Dayan, P. R. Montague, A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
2. Y. Niv, Reinforcement learning in the brain. *J. Math. Psychol.* **53**, 139–154 (2009).
3. V. de Lafuente, R. Romo, Dopamine neurons code subjective sensory experience and uncertainty of perceptual decisions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19767–19771 (2011).
4. K. Nomoto, W. Schultz, T. Watanabe, M. Sakagami, Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. *J. Neurosci.* **30**, 10692–10702 (2010).
5. A. Lak, K. Nomoto, M. Keramati, M. Sakagami, A. Kepecs, Midbrain dopamine neurons signal belief in choice accuracy during a perceptual decision. *Curr. Biol.* **27**, 821–832 (2017).
6. S. Sarno, V. de Lafuente, R. Romo, N. Parga, Dopamine reward prediction error signal codes the temporal evaluation of a perceptual decision report. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E10494–E10503 (2017).
7. R. P. N. Rao, Decision making under uncertainty: A neural model based on partially observable Markov decision processes. *Front. Comput. Neurosci.* **4**, 146 (2010).
8. C. K. Starkweather, B. M. Babayan, N. Uchida, S. J. Gershman, Dopamine reward prediction errors reflect hidden-state inference across time. *Nat. Neurosci.* **20**, 581–589 (2017).
9. B. M. Babayan, N. Uchida, S. J. Gershman, Belief state representation in the dopamine system. *Nat. Commun.* **9**, 1891 (2018).
10. S. J. Gershman, N. Uchida, Believing in dopamine. *Nat. Rev. Neurosci.* **20**, 703–714 (2019).
11. W. Schultz, Multiple dopamine functions at different time courses. *Annu. Rev. Neurosci.* **30**, 259–288 (2007).
12. T. Sawaguchi, P. S. Goldman-Rakic, D1 dopamine receptors in prefrontal cortex: Involvement in working memory. *Science* **251**, 947–950 (1991).
13. T. Satoh, S. Nakai, T. Sato, M. Kimura, Correlated coding of motivation and outcome of decision by dopamine neurons. *J. Neurosci.* **23**, 9913–9923 (2003).
14. R. Romo, C. D. Brody, A. Hernández, L. Lemus, Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* **399**, 470–473 (1999).
15. A. Akrami, C. D. Kopec, M. E. Diamond, C. D. Brody, Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature* **554**, 368–372 (2018).
16. P. Ashourian, Y. Loewenstein, Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS One* **6**, e19551 (2011).
17. A. Fassihi, A. Akrami, V. Esmaeili, M. E. Diamond, Tactile perception and working memory in rats and humans. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2331–2336 (2014).
18. C. Preuschhof, T. Schubert, A. Villringer, H. R. Heekeren, Prior information biases stimulus representations during vibrotactile decision making. *J. Cogn. Neurosci.* **22**, 875–887 (2010).
19. C. D. Brody, A. Hernández, A. Zainos, R. Romo, Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb. Cortex* **13**, 1196–1207 (2003).
20. R. Romo, R. Rossi-Pool, Turning touch into perception. *Neuron* **105**, 16–33 (2020).
21. M. Matsumoto, M. Takada, Distinct representations of cognitive and motivational signals in midbrain dopamine neurons. *Neuron* **79**, 1011–1024 (2013).
22. T. Ott, A. Nieder, Dopamine and cognitive control in prefrontal cortex. *Trends Cogn. Sci.* **23**, 213–234 (2019).
23. J. D. Berke, What does dopamine mean? *Nat. Neurosci.* **21**, 787–793 (2018).
24. A. A. Hamid *et al.*, Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* **19**, 117–126 (2016).
25. Y. Niv, N. D. Daw, D. Joel, P. Dayan, Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* **191**, 507–520 (2007).
26. M. Rothkirch, P. Sterzer, "The role of motivation in visual information processing" in *Motivation and Cognitive Control, 1/E*, T. S. Braver, Ed. (Routledge, 2015), . pp. 35–61.
27. W. Schultz, R. Romo, Responses of nigrostriatal dopamine neurons to high-intensity somatosensory stimulation in the anesthetized monkey. *J. Neurophysiol.* **57**, 201–217 (1987).
28. E. S. Bromberg-Martin, M. Matsumoto, O. Hikosaka, Distinct tonic and phasic anticipatory activity in lateral habenula and dopamine neurons. *Neuron* **67**, 144–155 (2010).
29. A. Kepecs, N. Uchida, H. A. Zariwala, Z. F. Mainen, Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455**, 227–231 (2008).
30. D. C. Knill, W. Richards, *Perception as Bayesian Inference* (Cambridge University Press, 1996).
31. W. J. Ma, M. Jazayeri, Neural coding of uncertainty and probability. *Annu. Rev. Neurosci.* **37**, 205–220 (2014).
32. R. Romo, A. Hernández, A. Zainos, L. Lemus, C. D. Brody, Neuronal correlates of decision-making in secondary somatosensory cortex. *Nat. Neurosci.* **5**, 1217–1225 (2002).
33. R. Rossi-Pool, A. Zainos, M. Alvarez, G. Diaz-deLeon, R. Romo, A continuum of invariant sensory and behavioral-context perceptual coding in secondary somatosensory cortex. *Nat. Commun.* **12** (2021).
34. R. Rossi-Pool, J. Vergara, R. Romo, "Constructing perceptual decision-making across cortex" in *The Cognitive Neurosciences, 6/E*, D. Poeppel, G. Mangun, M. Gazzaniga, Eds. (The MIT Press, 2020) 413, 427.
35. V. de Lafuente, R. Romo, Dopaminergic activity coincides with stimulus detection by the frontal lobe. *Neuroscience* **218**, 181–184 (2012).

Sarno et al.
Dopamine firing plays a dual role in coding reward prediction errors and signaling motivation in a working memory task

PNAS | 9 of 10
https://doi.org/10.1073/pnas.2113311119

NEUROSCIENCE

36. J. G. Mikhael, H. R. Kim, N. Uchida, S. J. Gershman, The role of state uncertainty in the dynamics of dopamine. bioRxiv [Preprint] (2021). https://doi.org/10.1101/805366. Accessed 11 August 2021.

37. R. Rossi-Pool *et al*., Emergence of an abstract categorical code enabling the discrimination of temporally structured tactile stimuli. *Proc. Natl. Acad. Sci. U.S.A* **113**, E7966–E7975 (2016).

38. R. Rossi-Pool *et al*., Temporal signals underlying a cognitive process in the dorsal premotor cortex. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7523–7532 (2019).

39. J. G. Mikhael, L. Lai, S. J. Gershman, *Rational Inattention and Tonic Dopamine. PLoS Comp. Biol.* **17**, e1008659 (2021).

40. T. Ljungberg, P. Apicella, W. Schultz, Responses of monkey midbrain dopamine neurons during delayed alternation performance. *Brain Res.* **567**, 337–341 (1991).

41. W. Schultz, P. Apicella, T. Ljungberg, Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* **13**, 900–913 (1993).

42. A. Mohebi *et al.*, Dissociable dopamine dynamics for learning and motivation. *Nature* **570**, 65–70 (2019).

43. M. W. Howe, P. L. Tierney, S. G. Sandberg, P. E. M. Phillips, A. M. Graybiel, Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* **500**, 575–579 (2013).

44. H. R. Kim *et al.*, A unified framework for dopamine signals across timescales. *Cell* **183**, 1600–1616.e25 (2020).

45. A. Westbrook, T. S. Braver, Dopamine does double duty in motivating cognitive effort. *Neuron* **89**, 695–710 (2016).

46. A. E. Hamilos *et al.*, Slowly evolving dopaminergic activity modulates the moment-to-moment probability of movement initiation. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2020.05.13.094904 (Accessed 5 May 2021).

47. A. Kato, K. Morita, Forgetting in reinforcement learning links sustained dopamine signals to motivation. *PLoS Comput. Biol.* **12**, e1005145 (2016).

48. K. Lloyd, P. Dayan, Tamping ramping: Algorithmic, implementational, and computational explanations of phasic dopamine signals in the accumbens. *PLoS Comput. Biol.* **11**, e1004622 (2015).

49. D. M. Yee, T. S. Braver, Interactions of motivation and cognitive control. *Curr. Opin. Behav. Sci.* **19**, 83–90 (2018).

50. W. Kool, M. Botvinick, The intrinsic cost of cognitive control. *Behav. Brain Sci.* **36**, 697–698, discussion 707–726 (2013).

51. A. Shenhav, M. M. Botvinick, J. D. Cohen, The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240 (2013).

52. M. Botvinick, T. Braver, Motivation and cognitive control: From behavior to neural mechanism. *Annu. Rev. Psychol.* **66**, 83–113 (2015).

53. R. Cools, The costs and benefits of brain dopamine for cognitive control. *Wiley Interdiscip. Rev. Cogn. Sci.* **7**, 317–329 (2016).

54. N. Brunel, X.-J. Wang, Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci.* **11**, 63–85 (2001).

55. S. Sarno, J. Falcó-Roget, MidbrainDopaFiringMotivationWM. Open Science Framework. 10.17605/OSF.IO/5X8DY. Deposited 16 November 2021.

10 of 10 | **PNAS**
https://doi.org/10.1073/pnas.2113311119

Sarno et al.
Dopamine firing plays a dual role in coding reward prediction errors and signaling motivation in a working memory task