Supplementary Information for:

Double dissociation of value computations in orbitofrontal and anterior cingulate neurons

Steven W. Kennerley, Timothy E. J. Behrens & Jonathan D. Wallis
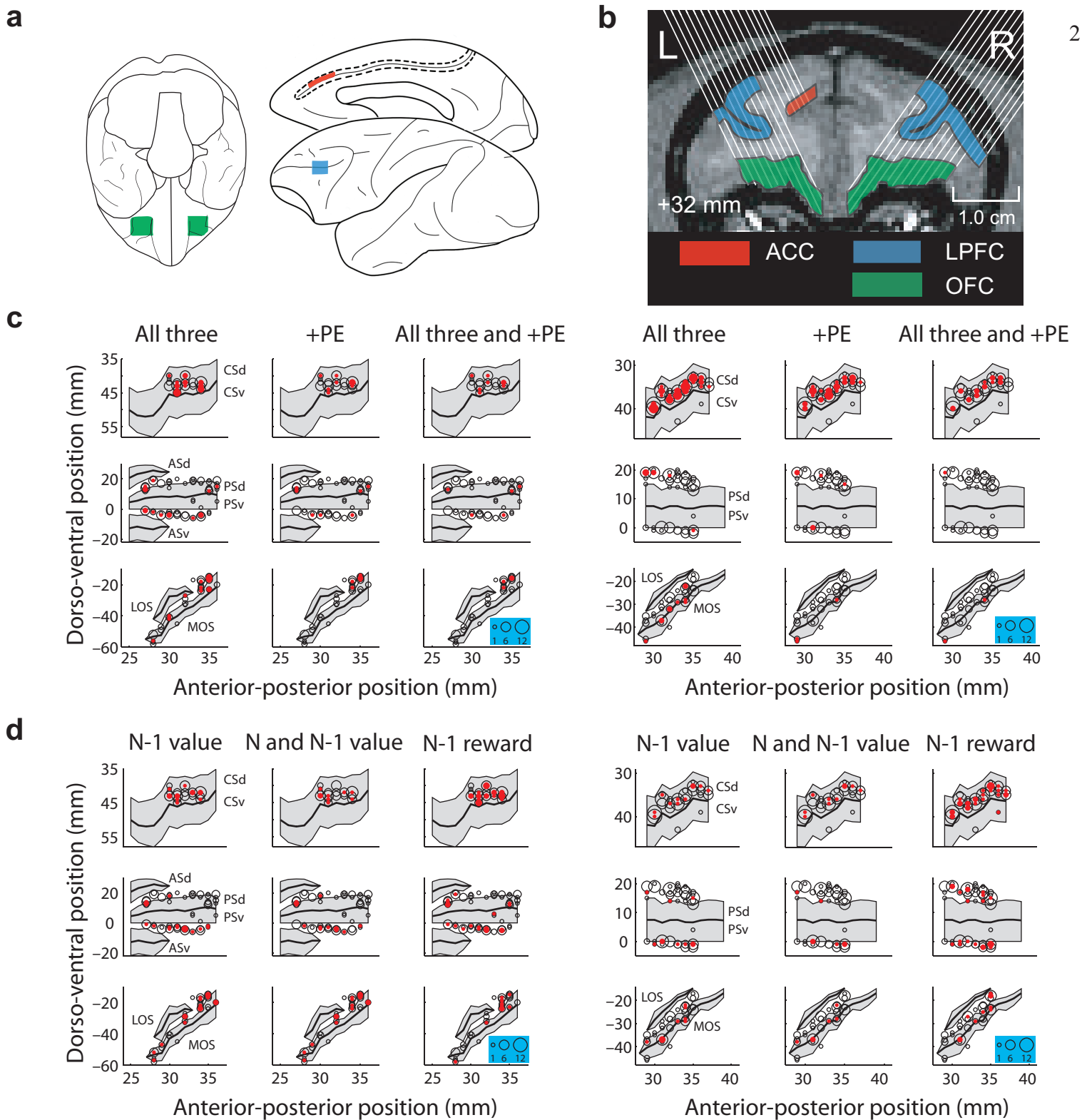
Content list:

Supplementary Figure 1. PFC recording locations. **(a)** Schematic showing the frontal boundaries on ventral, medial and lateral views of the macaque brain. Dashed line depicts unfolded cingulate sulcus. **(b)** Magnetic resonance scan of a coronal slice through the frontal lobe of subject B. Shaded regions denote the boundaries of the three frontal areas investigated. White lines depict potential electrode paths. The coronal slice is at the approximate center position of the recording chambers along the anterior-posterior axis. **(c)** Locations of all recorded neurons (open circles) and neurons selective (filled circles) for encoding all three decision variables (All three), positive prediction error (+PE) or both (All three and +PE) in subject A (left) and subject B (right). **(d)** Same as in (c) but neurons selective for previous trial overall value (N–1), previous and current trial overall value (N and N–1) and previous trial reward value. The top, middle and bottom panels correspond to areas ACC, LPFC and OFC, respectively. We measured the anterior-posterior position from the interaural line (x-axis), and the dorso-ventral position relative to the lip of the ventral bank of the principal sulcus (0 point on y-axis). Gray shading indicates unfolded sulci. CSd/CSv = dorsal/ventral bank of the anterior cingulate sulcus, ASd/ASv = dorsal/ventral spur of arcuate sulcus, PSd/PSv = dorsal/ventral bank of principal sulcus, LOS/MOS = lateral/medial orbital sulcus.
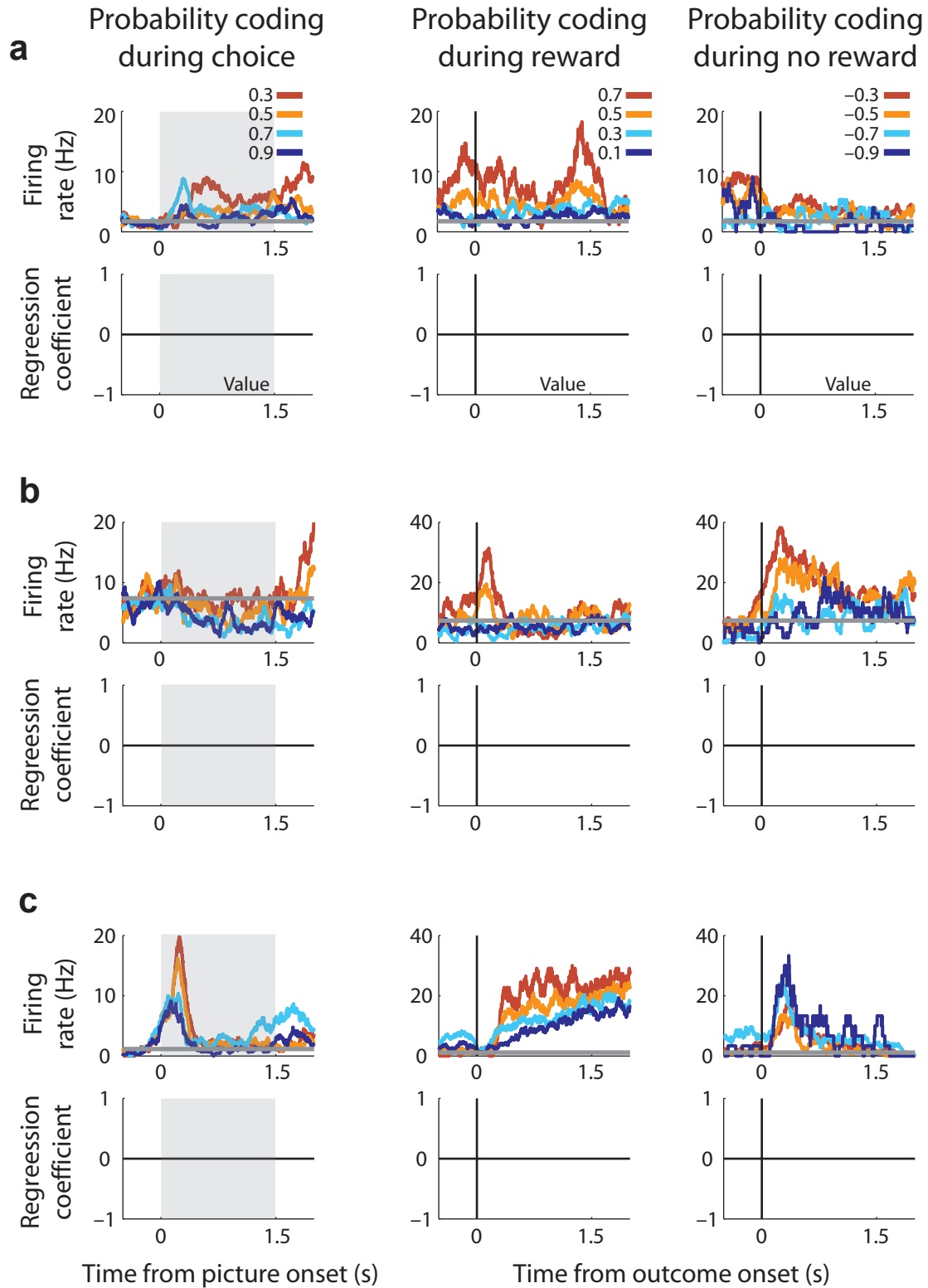
Supplementary Figure 2. Three single neurons showing activity on probability trials during both the choice and outcome epochs. Conventions as in Figure 2. **(a)** An LPFC neuron that encoded reward probability with a negative regression coefficient during the choice epoch which was maintained during the outcome epoch, reflecting a value signal at both choice and outcome. **(b)** An ACC neuron that encoded reward probability at choice with a negative regression coefficient, and strongly encoded reward probability on both rewarded and unrewarded outcomes also with a negative regression coefficient, thus reflecting a value signal at both choice and outcome. **(c)** An ACC neuron that encoded a salience signal. This neuron encoded reward probability during both rewarded and unrewarded outcomes but with opposite signed regression coefficients. In other words, this neuron increased firing rate as the absolute value of the prediction error increased, suggesting it encoded the discrepancy between expected and experienced outcome but in an unsigned manner.

Reinforcement history analysis

Overall value history analysis

We examined whether our population of PFC neurons encoded the overall value history of choices as well as the reward amount received on the last trial (GLM-3; see Methods). This analysis was performed across all trials with a focus on three different task epochs; during the 1000-ms fixation period prior to picture onset, during the 1500-ms choice epoch, and during the 1000-ms outcome epoch of the current trial. The logic here was that we wanted to investigate when the value information about the previous trial first arose (i.e., did it carry over from the previous trial) and when it decayed.

We first determined the number of neurons in each PFC area which encoded the past trial overall value during these three different epochs (Supplementary Figure 3a–c). At the time of fixation prior to the current choice, there was no difference between the number of neurons across PFC areas which encoded N–1 overall value (Supplementary Fig. 3a; χ^2=3.3, P>0.1). Yet as soon as the pictures were presented and the current choice could be evaluated, significantly more neurons in OFC encoded N–1 overall value (Supplementary Fig. 3b; χ^2=6.2, P=0.04). OFC neurons continued to significantly encode N–1 overall value into the outcome epoch of the current choice, but only when compared to LPFC neurons (Supplementary Fig. 3c; χ^2=8.7, P=0.01). Thus, OFC neurons do not simply encode the previous trial value as a working memory signal that persists from the time of the previous outcome. Instead, the value history representation only emerges at the time of the current choice, as if this value representation is used as a reference for which to base the encoding of the current choice value.

In the value history analysis in the main text, we found that the activity of all OFC neurons exhibited a strong correlation between the current choice overall value and the previous history of choice values (Fig. 8g–h). In other words, as a population, OFC neurons simultaneously encode information about the current and past overall choice value, emerging around 300-ms into the current choice evaluation. This correlation was negative, indicating that the firing rate of the current choice increased as the previous trial value decreased (i.e., as the discrepancy between the value of the current and past choice values increased).

Most of our analyses require neurons to reach a specific level of significance before they are classified as encoding a specific value parameter (see Methods). This ensures that our Type 1 error levels were appropriate. However, it is possible that this selectivity criterion also excludes some neurons that might be encoding a particular value parameter but that do not quite reach our level of significance (Type II error). This is a particularly important issue for the reinforcement history analysis for two reasons. First, the correlation analyses shown in Figure 8g–h requires neurons to reach significance for two independent value representations (i.e., current and past trial value), yet we did not relax our selectivity criterion for reaching this classification. Second, despite relatively few neurons reaching significance for both of these value parameters (Fig. 8f), the correlation between these two value parameters within the entire OFC population was very high. This suggested that this correlation might be at least partially influenced by neurons that did not quite reach our selectivity criterion for one of the two value parameters.

To examine this issue directly, we excluded all of the selective PFC neurons that encoded both the current and past trial overall value (Fig. 8f) and performed this correlation again. This excluded a total of 21 (8%), 23 (16%) and 19 (9%) of LPFC, OFC and ACC neurons, respectively. While this exclusion procedure removed any correlation between current and past overall value coding in the firing rates of LPFC and ACC neurons, the neuronal activity within the remaining OFC population exhibited a very strong anti-correlation between current overall value and the overall choice value from one (Supplementary Fig. 3d) and two (Supplementary Fig. 3e) trials into the past. This implies that as a population, OFC neurons encode an adaptive choice value signal that is dynamically adapted based on the recent history of choice values, whereas neurons in both ACC and LPFC encode current choice value in a way that is insensitive to past choice values.

Reward history analysis

We next determined the number of neurons in each PFC area which encoded the past trial reward value (magnitude) at the time of the current trial's fixation, choice and outcome epochs, respectively (Supplementary Fig. 3f–h). At the time of fixation prior to the current choice, many neurons in ACC encoded the amount of reward received on the previous trial (Supplementary Fig. 3f; $\chi^2=33$, $P<9\text{x}10^{-8}$). Although this signal could reflect information about the magnitude of liquid reward still present in the mouth, it is noteworthy that the fixation epoch occurs after a minimum of 2.5s (ITI and trial termination delays) from the termination of reward delivery, suggesting that this signal may instead reflect a reward working memory signal[1]. There were also significantly more ACC neurons which encoded the reward value of the previous trial at the time of the current choice (Supplementary Fig. 3g; $\chi^2=6.0$, $P=0.05$), and there was a significant tendency for these ACC neurons to have a –RC (34/52; z-score=2.4, $P=0.018$ binomial test). No significant trial N–1 reward effects were present in the outcome epoch of the current trial (Supplementary Fig. 3h).
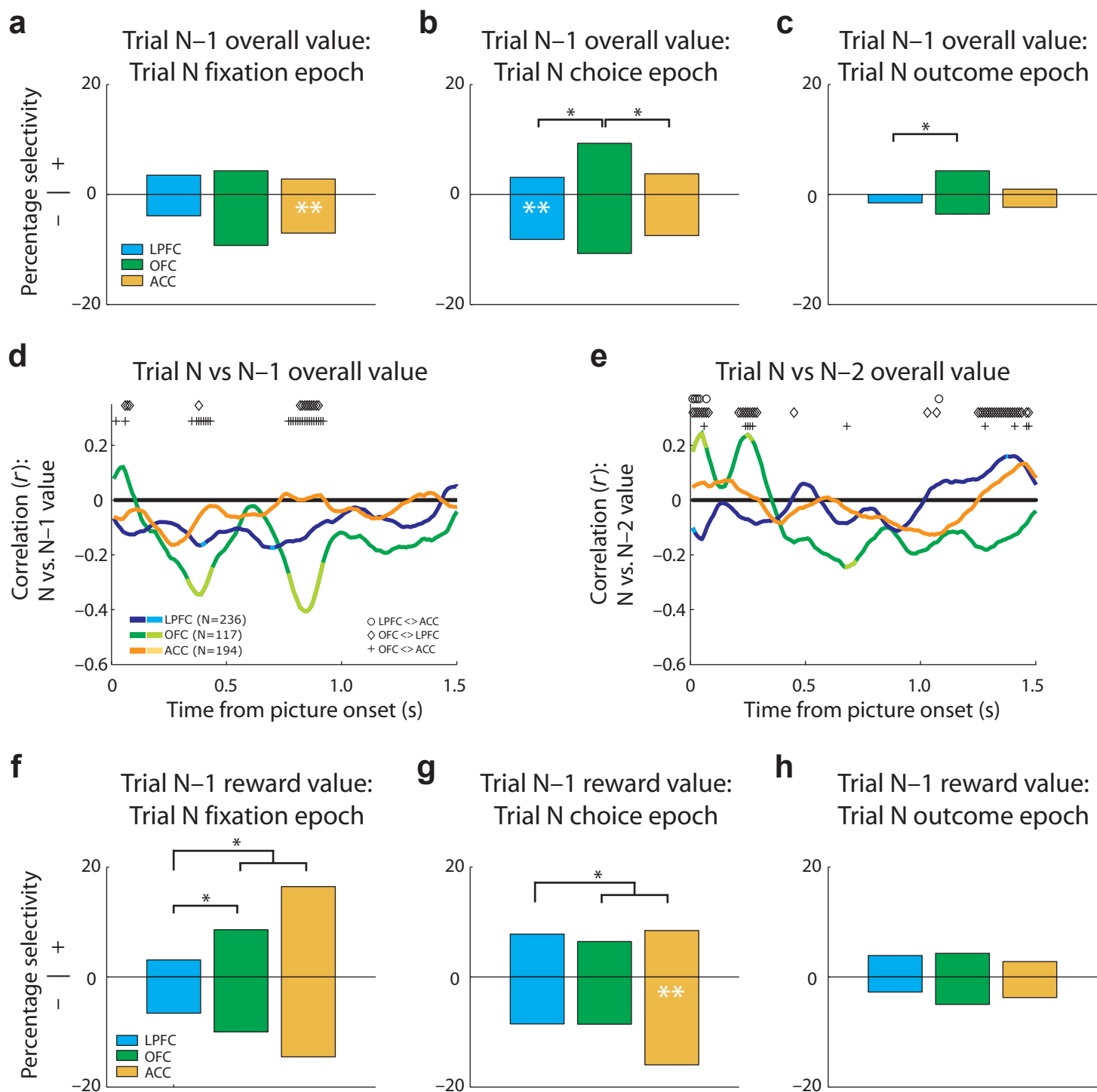
Supplementary Figure 3. Neuronal encoding of value history and reward history. **(a–c)** Prevalence of neurons encoding the overall value of the previous (N–1) trial at the time of fixation **(a)**, choice **(b)**, or outcome **(c)** of the current trial (N) with a positive (+) or negative (–) regression coefficient. Conventions as in Figure 3. OFC neurons were significantly more likely to encode the overall value of the past trial compared to ACC and LPFC, but only at the time the current choice was assessed. Thus, this value signal emerged only once the current choice could be evaluated. **(d–e)** Mean correlation between encoding of the overall value of the current and N–1 **(d)** or N–2 **(e)** trial but when all of the significant neurons from Figure 8f are removed. Conventions as in Figure 8. The anti-correlation between present and past overall value was evident even in the non-significant population of OFC neurons. **(f–h)** Prevalence of neurons encoding the reward value of the previous (N–1) trial at the time of fixation **(f)**, choice **(g)**, or outcome **(h)** of the current trial with a positive or negative regression coefficient. ACC neurons were significantly more likely to encode the reward amount received on the past trial at the time of fixation or choice of the current trial. Thus, ACC neurons appeared to maintain a representation of the previous reward history in working memory that carried into the next trial.

Behavioral adaptation analysis

To examine adaptive behavior following errors or unrewarded responses, we collapsed across the side presentation of stimuli such that we examined choice behavior for three decision variables, four choice values and two picture sets (see Fig. 1b). We calculated the percent correct per condition across the 24 and 35 behavioral sessions in which neuronal data was collected in subjects A and B, respectively (Supplementary Fig. 4a). Across all conditions, the two subjects performed a total of 32,963 trials with only 626 errors (98.1% correct). Subjects had error rates of <3% on 19/24 conditions, with 177/626 errors (28% of all errors) occurring on the two conditions where the 0.7 stimulus from picture set 2 was presented. Apart from these two conditions, subjects made one or fewer errors per condition in >80% of all behavioral sessions. When subjects made at least one error within a condition, we determined the likelihood of making a correct choice the next time the same condition was presented within the behavioral session (i.e., error correction). Across all conditions and sessions, subject A corrected 215/253 (85%; z-score=11.1, $P<1\times10^{-20}$ binomial test) and subject B corrected 241/281 (86%; z-score=12.0, $P<1\times10^{-20}$ binomial test) of their errors. Thus, both subjects rarely made errors and when they did, they were significantly above chance at correcting their error on the next occurrence of the same condition. The lack of error trials across conditions within a behavioral session yielded insufficient power for any neuronal analysis of error-related selectivity.

To explore whether subjects adapted their future choices based on whether current choices were rewarded, we sorted all probability trials within a behavioral session by condition and examined successive pairs of trials of the same condition (trials N and N+1). For example, the 0.1 vs. 0.3 probability condition might occur on trial 9, 28, 81, etc., and our analysis would assign trial 9 as "trial N" and examine whether trial 28 ("trial N+1") was correct, and then repeat the same procedure for trials 28 and 81 and so forth until all trials within a condition were examined. The logic for this analysis was that if subjects were basing their choices on the history of probabilistic outcomes (hence adaptive decision-making) rather than the well-learned associations between stimuli and outcomes then they should be less likely to select correctly on trial N+1 if trial N was a correct choice that was not rewarded.

We determined performance on probability trial N+1 when each subject chose correctly on probability trial N (Supplementary Fig. 4b). This analysis revealed that subjects were relatively indifferent to the outcome (i.e., reward or no-reward) of a correct choice on trial N; regardless of whether or not a correct choice was rewarded on trial N, subjects chose correctly on over 90% of choices when the same condition was next presented (trial N+1). For example, despite the fact that subjects received no reward on ~70% of trials following a choice of the 0.3 probability stimulus in the 0.1 vs. 0.3 condition (value=1, red bars), subjects only changed their behavior (and hence selected the incorrect 0.1 probability stimulus) on trial N+1 <2% of the time. Thus we found no evidence that subjects were more likely to adapt their behavior on trial N+1 (i.e., select the less optimal stimulus) if a correct trial N choice was not rewarded. Collectively, these results suggest that the choices of both subjects were guided primarily by well learned predictions between a stimulus and its associated value rather than being influenced by the session history of choice outcomes.
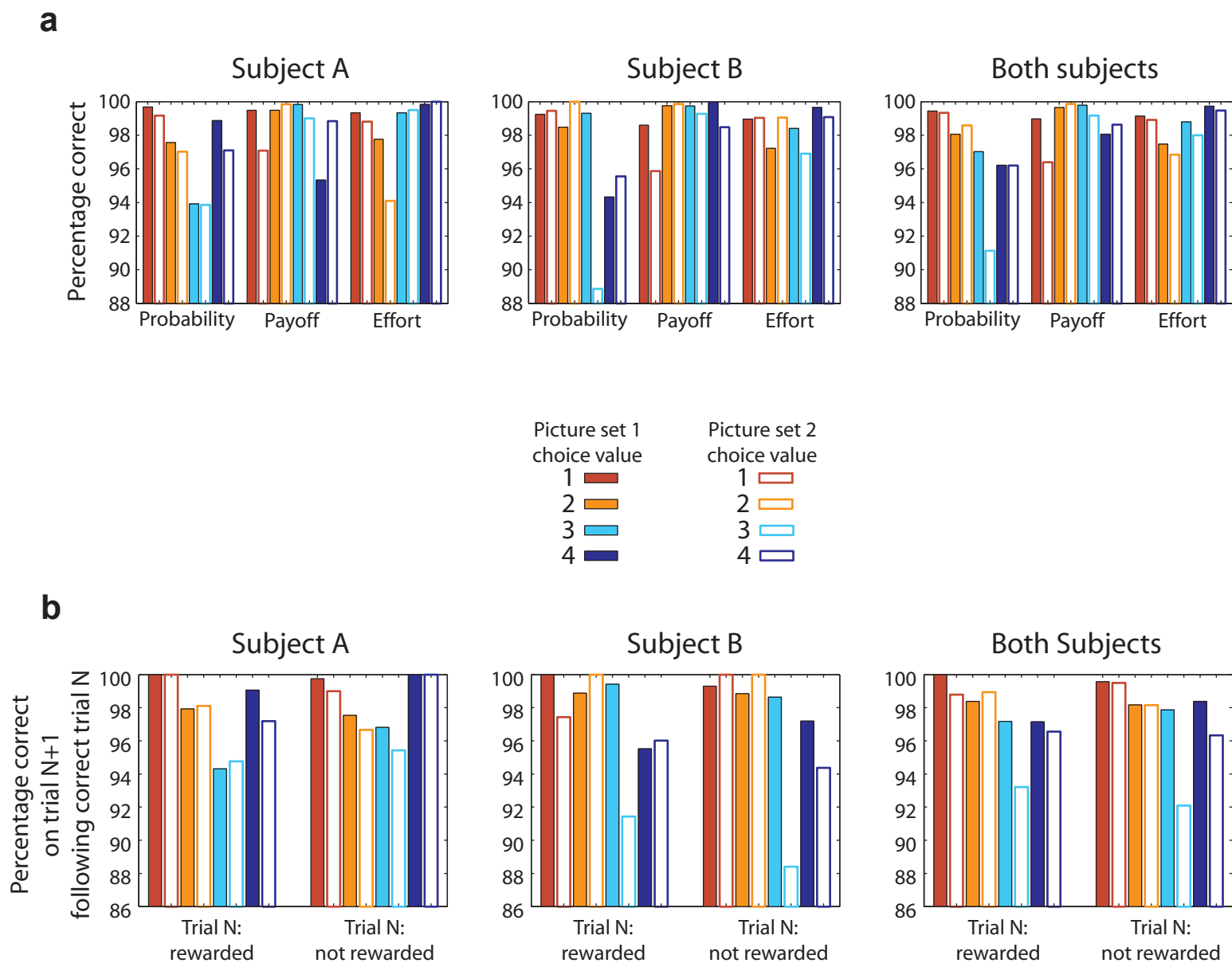
Supplementary Figure 4. Percentage correct choices per decision variable or reward history. **(a)** Percentage correct per choice condition defined as the percentage of completed trials where the more valuable option was chosen. Choice value ranges from 1–4 (see Fig. 1b), where a value of 1 is the lowest value for any decision variable (e.g., 0.1 vs. 0.3 probability choice) and a value of 4 is the most valuable for any decision variable (e.g., 0.7 vs. 0.9 probability choice). Filled and unfilled bars indicate performance on choices from picture sets 1 and 2, respectively. **(b)** Future performance across the different probability conditions as a function of whether past choices were rewarded. For each probability condition when a correct response was made on trial N, the plot shows the percentage correct on the trial when the same condition was next presented (trial N+1) as a function of choice outcome (reward versus no-reward) on trial N. Whether or not a subject was rewarded for a correct response on trial N had virtually no influence on whether they would choose correctly on the same condition when next presented (trial N+1).

References

51.   Lara, A. H., Kennerley, S. W. & Wallis, J. D. Encoding of gustatory working memory by orbitofrontal neurons. *J Neurosci* **29**, 765-74 (2009).