Cerebellar climbing fibers encode expected reward size

Noga Larry[1*], Merav Yarkoni[1*], Adi Lixenberg[1] and Mati Joshua[1]

1. Edmond and Lily Safra Center for Brain Sciences, the Hebrew University, Jerusalem, Israel

   * These authors contributed equally.

Climbing fiber inputs to the cerebellum encode error signals that instruct learning. Recently, evidence has accumulated to suggest that the cerebellum is also involved in the processing of reward. To study how rewarding events are encoded, we recorded the activity of climbing fibers when monkeys were engaged in an eye movement task. At the beginning of each trial, the monkeys were cued the size of the reward that would be delivered upon successful completion of the trial. We found increased climbing fiber activity during cue presentation when information about reward size was first made available. Reward size did not modulate activity at reward delivery. These results indicate that climbing fibers encode the expected reward size and suggest a general role of the cerebellum in associative learning beyond error correction.

Introduction

There is strong computational, anatomical, and functional evidence that support the theory that the cerebellar cortex performs error correcting supervised motor learning (Albus, 1971; Gilbert and Thach, 1977; Marr, 1969; Nguyen-Vu et al., 2013b; Stone and Lisberger, 1990; Suvrathan et al., 2016). In this framework, motor learning occurs through changes in the computation of Purkinje cells, the sole output cells of the cerebellar cortex. Purkinje cells receive two distinct types of inputs: parallel fiber and climbing fiber. Each type of input leads to a different type of action potential. Parallel fiber inputs modulate the rate of simple spikes, events similar to action potentials in other cell types. Climbing fiber inputs lead to complex spikes (CS), which are unique prolonged events. CS are thought to represent instructive error signals triggered by movement errors. These error signals adjust the simple spike response of the Purkinje cell to parallel fiber input, leading to improvement in subsequent movements. The hypothesized role of the CS in learning was broadened when it was shown that the CS rate increases in response to cues that are predictive of undesired successive stimuli (Ohmae and Medina, 2015). Thus, the CS signal is well-suited for driving associative learning based on avoidance of aversive stimuli.

Recent research has shown that CS rate increases when behavior leads to a desired rewarded outcome (Heffley et al., 2018) , a marked departure from their established role in error signaling. The CS reward signal could be directly linked to reward consumption behavior, such as licking (Welsh et al., 1995) or to the signal at reward delivery that behavior was successful (Heffley et al., 2018). Alternatively, the CS could encode the predicted reward consequences of arbitrary stimuli, similar to the way in which CS encode the prediction of an undesired air-puff (Ohmae and Medina, 2015). The critical question is thus whether CS increase to reward predictive stimuli. To probe this issue, this study examined CS responses to reward predictive cues.

We designed a task that temporally separated reward information, motor behavior and reward delivery (Joshua and Lisberger, 2012). We found that climbing fiber activity encoded the expected reward size seconds before the reward delivery. Reward size did not modulate activity at reward delivery. These findings imply that the cerebellum receives signals that could allow it to perform both error and reward-based associative learning, thus going beyond the accepted role of the cerebellum in error correction to suggest a general role in associative learning.

We additionally studied the effect of reward expectation on CS coding during eye movements. We found that reward expectation did not modulate the CS tuning of movement parameters. During the cue, the CS and simple spike rates of cells were uncorrelated, in contrast to the negative correlation that has been reported in the context of error correction learning. This suggests that CS can instruct behavioral change through different mechanisms.

Results

Complex spikes encode the size of the expected reward

We recorded climbing fiber activity while monkeys performed a smooth pursuit eye movement task in which we manipulated the expected reward size (Joshua and Lisberger, 2012; Fig. 1A). At the start of each trial, the monkey fixated on a white spot. The spot then changed to one of two colors, indicating whether a large or small reward would be given upon successful completion of the trial. After a variable delay, the colored target began to move in one of eight directions and the monkey had to accurately track it. At the end of a successful trial, the monkey received either a large or a small reward, as indicated by the color of the cue.

During the taskmonitored eye movements and recorded neural activity from the ventral parts of the cerebellum (Supplementary Fig. 1). Our recordings included neurons that responded to eye movements (Stone and Lisberger, 1990) and neurons that did not. Our task design allowed us to separately analyze the CS rate following cue presentation, during pursuit, and following reward delivery. The average eye velocity during tracking of the large reward target was faster and more similar to the target velocity, compared to tracking of the small reward target (Fig. 1B). The difference was evident even at the single session level. In most sessions, the average eye velocity 250ms following motion onset was larger when the expected reward was large (Fig. 1C). This behavioral difference and the selection of the larger reward target in an additional choice task (Supplementary Fig. 2) indicate that the monkeys associated the reward size with the color of the target.

Following the presentation of the cue, we found many Purkinje cells that transiently increased their CS rate when the expected reward was large but not when the expected reward was small (examples in Fig. 1D and Supplementary Fig. 3). This difference was apparent when examining the population average CS peri-stimulus time histogram (PSTH). After the color cue appeared, the population average CS rate was higher when the expected reward was large, as can be seen by the difference in the PSTHs of the two reward conditions (Fig.

1E). At the single cell level, most cells had a higher CS rate on large reward trials than in small reward trials (Fig. 1F, most dots lie beneath the identity line). Thus, the CS rate is modulated by changes in reward expectation, at times temporally distinct from the behavioral effect on pursuit eye movements and reward delivery.

Complex spikes do not encode reward size at reward delivery

The population CS rate was only affected by reward size when information regarding future reward was given, but not during the reward itself. During reward delivery, the PSTHs of the two conditions overlapped (Fig. 2A), indicating a similar population response for the large and small rewards. When examining the responses of single cells, the CS rate was similar in the two reward conditions in that most cells fell close to the identity line (Fig. 2B). To compare the temporal pattern of the reward size encoding at cue and reward delivery, we calculated the difference in PSTHs between the large and small reward conditions (Fig. 2C). The difference between large and small reward rose sharply shortly after the color cue appeared. In sharp contrast, following reward delivery we found only a small rate fluctuation that resembled the fluctuation prior to reward delivery. At the single cell level there was no correlation between cell encoding of reward size during the cue and during reward delivery. For both the full population and for the subpopulation of neurons significantly coding the reward size at cue, the correlation between cue and reward delivery epochs was not significant (Fig.2D). This indicates that Purkinje cells that differentiated reward conditions during the cue did not differentiate between them during delivery.

We ruled out the possibility that differences in licking behavior was responsible for the CS rate modulations. The pattern of licking (Fig. 2E,F) and CS rate modulation was completely different. Licking but not spiking increased at reward delivery. Further, after cue onset, licking in both reward conditions decreased whereas the temporal pattern of CS was different between reward conditions (Fig. 1E). We also analyzed the pattern of saccades and microsaccades and found that it also differed from the CS pattern (Supplementary Fig. 4).

Complex spike tuning to the direction of motion does not depend on reward size

Overall, these results indicate that reward expectation, but not concurrent behavior, affects the CS rate. However, CS can also signal movement initiation in a directionally tuned manner (Kobayashi et al., 1998; Stone and Lisberger, 1990). To determine how CS coding of eye movement parameters is affected by reward expectation, we identified Purkinje cells that were directionally tuned (examples in Fig. 3A and Supplementary Fig. 5). When we examined

the CS rate in the preferred direction of the cell and the direction 180° to it (the null direction), we did not find differences in the CS rate between reward conditions (Fig. 3B). We also aligned cells to their PD and calculated a population tuning curve for each reward condition. The tuning curves overlapped and were not significantly different (Fig. 3C). Moreover, when we extended the task and tested for modulation at three different target speeds, there was no difference between reward conditions (Supplementary Fig. 6). Thus, encoding of reward is limited to the time point at which the reward size is first signaled and not the time when reward drives changes in behavior.

The directionally tuned CS signal has been linked to the coding of visual errors that instruct motor learning (Medina and Lisberger, 2008; Nguyen-Vu et al., 2013a) by changing the simple spike response to parallel fiber inputs. Our result indicate that this signal is not modulated by reward expectation. A possible intriguing population organization is that Purkinje cells in which the CS encode motor errors constitute a different subpopulation from Purkinje cells in which CS encode reward expectation. In our data, we were unable to detect such organization. Although coding of reward size during the cue was slightly smaller for the directionally tuned CS, the difference in the encoding of reward between directionally tuned and non-tuned CS was not significant (Supplementary Fig. 7A-C). We also examined the coding of the cued reward size in cells that were directionally tuned in their simple spike responses and cells that were not but the differences did not reach significance (Supplementary Fig. 7D-F).

The relation between simple and complex spikes is different for reward and direction tuning

CS generate plasticity in parallel fiber synapses leading to a decrease in the simple spike rate (Ekerot and Kano, 1985). This plasticity is thought to underlie the opposite modulations of simple and CS rates on different tasks (Badura et al., 2013; Gilbert and Thach, 1977; Stone and Lisberger, 1990). In our dataset, we saw similar modulations during movement. We aligned the CS tuning curve to the preferred direction of the simple spikes of the same cell. This revealed that the CS rate decreased in directions for which the simple spike rate increased (Fig. 3D). To examine whether this effect existed at the single cell level, we calculated the signal correlation of complex and simple spikes which we defined as the correlation between simple and complex direction tuning curves. We found that most signal correlations were negative; in other words that the CS and simple spikes were oppositely modulated during movement in most cells (Fig. 3E). This effect disappeared when we shuffled

the phase of the CS tuning curve or assigned direction labels randomly (see experimental procedures).

Unlike movement related modulation, the complex and simple spike reward modulations following cue presentation were not oppositely modulated (Fig. 3F). If reward-related modulations in CS drive simple spike attenuation, we would expect that the higher CS rate in the large reward condition would result in a stronger attenuation of simple spikes. This would lead to a negative correlation between the complex and simple spike reward modulations during the cue. However, we found that simple and CS modulations following cue presentation were uncorrelated (Fig. 3F). Further, the correlation was not significantly different from zero both whether we analyzed the full population or only those cells that significantly were tuned to reward size during the cue. Thus, the way the difference in CS rate during cue affects simple spike encoding and behavior may differ from the one suggested by the error signal model (Rowan et al., 2018).

Discussion

The difference in CS rate during cue presentation and the lack of difference during reward delivery strongly imply that CS can act as a reward predictions signal. This finding diverges from the accepted error signal model. The coding of predictive stimuli has been reported in CS in the context of error-based learning (Ohmae and Medina, 2015). Together with the current results, this suggests a more general role for the cerebellum in associative learning, when learning is both error and reward-based (Thoma et al., 2008; Wagner et al., 2017).

Similar reward signals have been found in dopaminergic neurons (Schultz et al., 1997), for which a time-difference learning model (Sutton and Barto, 2005) was suggested. A time difference learning model would imply that CS encode changes in expected future reward. This model makes strong predictions regarding the CS rate when reward is predicted with high certainty but omitted, or predicted with low certainty and delivered. The relationship between the certainty of reward prediction and CS rate requires further investigation in the future.

Error based models of the cerebellum link between the cerebellar representation of movement, plasticity mechanisms and learning. In this framework, the behavioral command of the cerebellar cortex in response to a stimulus is represented by the simple spike rate of Purkinje cells. CS lead to a reduction in the synaptic weight in recently active parallel fibers

and thereby change the simple spike rate in response to similar parallel fiber input. This change in the simple spike rate is hypothesized to alter the behavioral response to the same stimulus. Thus, when errors occur, the behavior that led to them is eliminated. The same logic cannot apply to learning from rewards since reward strengthens rather than eliminates the behavior that led to the reward (Thorndike, 1898). Indeed, reward-related modulation of CS did not exhibit the classical decrease in simple spike activity associated with CS activity. Parallel fiber plasticity could be modulated to allow reinforcement of the behavior that led to maximizing the upcoming reward (Rowan et al., 2018).

A further demonstration for independent mechanisms for learning from reward and sensory errors emerges when combining the current results with our recent behavioral study (Joshua and Lisberger, 2012). In that study, monkeys learned to predict a change in the direction of target motion by generating predictive pursuit movements. The size of the reward did not modulate learning process itself but only the execution of the movement (Joshua and Lisberger, 2012). The critical signal for direction change learning has been shown to be the directionally tuned CS signal (Medina and Lisberger, 2008). Our findings that the direction signal is not modulated by reward provides a plausible explanation at the implementation level for this behavioral finding. The directionally tuned CS that drive learning are not modulated by reward; therefore, learning itself is reward independent.

The current study demonstrates how climbing fibers encode predicted reward size. The reward signal is not limited to the direct rewarding consequences of the behavior. Thus we have found signals that can be used by the cerebellum to drive behavior that maximizes upcoming reward, enriching the signals that were previously found to eliminate undesired behaviors.

References

Albus, J.S. (1971). A theory of cerebellar function. Math. Biosci. *10*, 25–61.

Badura, A., Schonewille, M., Voges, K., Galliano, E., Renier, N., Gao, Z., Witter, L., Hoebeek, F.E., Chédotal, A., and DeZeeuw, C.I. (2013). Climbing fiber input shapes reciprocity of purkinje cell firing. Neuron *78*, 700–713.

Ekerot, C.F., and Kano, M. (1985). Long-term depression of parallel fibre synapses following stimulation of climbing fibres. Brain Res. *342*, 357–360.

Gilbert, P.F.C., and Thach, W.T. (1977). Purkinje cell activity during motor learning. Brain Res. *128*, 309–328.

Heffley, W., Song, E.Y., Xu, Z., Taylor, B., Hughes, M.A., McKinney, A., Joshua, M., and Hull, C. (2018). Coordinated cerebellar climbing fiber activity is gated by behavioral context in a voluntary motor regime. BioRxiv *21*, 281055.

Joshua, M., and Lisberger, S.G. (2012). Reward Action in the Initiation of Smooth Pursuit Eye Movements. J. Neurosci. *32*, 2856–2867.

Kobayashi, Y., Kawano, K., Takemura, A., Inoue, Y., Kitama, T., Gomi, H., and Kawato, M. (1998). Temporal Firing Patterns of Purkinje Cells in the Cerebellar Ventral Paraflocculus During Ocular Following Responses in Monkeys II. Complex Spikes. J. Neurophysiol. *80*, 832–848.

Marr, D. (1969). A Theory of Cerebellar Cortex. J. Physiol. 437–470.

Medina, J.F., and Lisberger, S.G. (2008). Links from complex spikes to local plasticity and motor learning in the cerebellum of awake-behaving monkeys. Nat. Neurosci. *11*, 1185–1192.

Nguyen-Vu, T.D., Kimpo, R.R., Rinaldi, J.M., Kohli, A., Zeng, H., Deisseroth, K., and Raymond, J.L. (2013a). Cerebellar Purkinje cell activity drives motor learning. Nat Neurosci *16*, 1734–1736.

Nguyen-Vu, T.D.B., Kimpo, R.R., Rinaldi, J.M., Kohli, A., Zeng, H., Deisseroth, K., and Raymond, J.L. (2013b). Cerebellar Purkinje cell activity drives motor learning. Nat. Neurosci. *16*, 1734–1736.

Ohmae, S., and Medina, J.F. (2015). Climbing fibers encode a temporal-difference prediction error during cerebellar learning in mice. Nat. Neurosci. *18*, 1798–1803.

Rowan, M.J.M., Bonnan, A., Zhang, K., Amat, S.B., Kikuchi, C., Taniguchi, H., Augustine, G.J., and Christie, J.M. (2018). Graded Control of Climbing-Fiber-Mediated Plasticity and Learning by Inhibition in the Cerebellum. Neuron *99*, 999–1015.e6.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science (80-.). *275*, 1593–1599.

Stone, L.S., and Lisberger, S.G. (1990). Visual responses of Purkinje cells in the cerebellar flocculus during smooth-pursuit eye movements in monkeys. II. Complex spikes. J. Neurophysiol. *63*, 1262–1275.

Sutton, R.S., and Barto, A.G. (2005). Introduction to Reinforcement Learning. Learning.

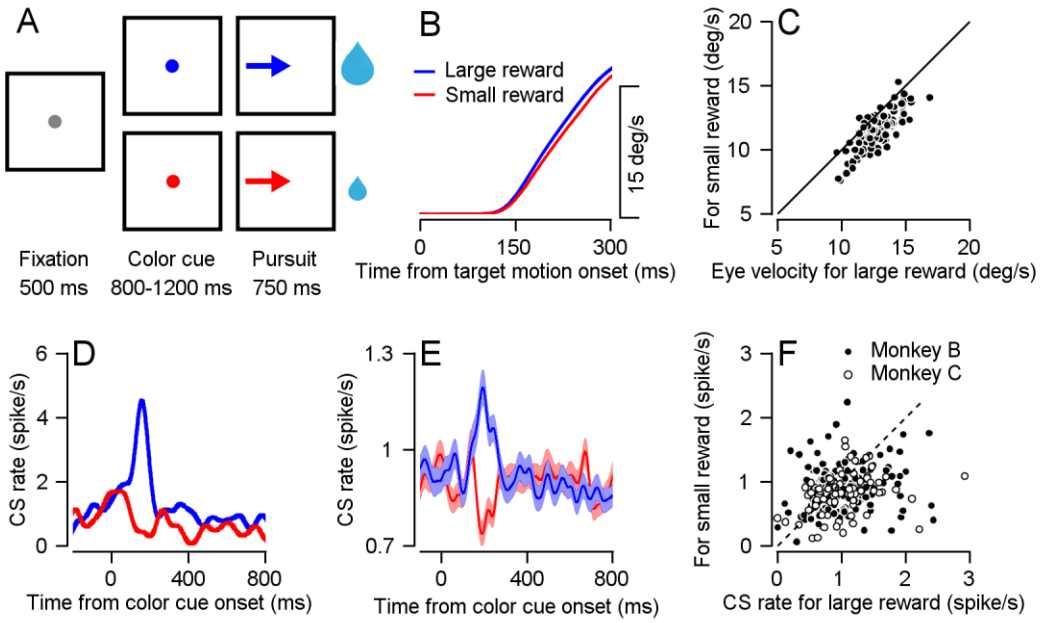
Suvrathan, A., Payne, H.L., and Raymond, J.L. (2016). Timing Rules for Synaptic Plasticity Matched to Behavioral Function. Neuron *92*, 959–967.

Thoma, P., Bellebaum, C., Koch, B., Schwarz, M., and Daum, I. (2008). The cerebellum is involved in reward-based reversal learning. Cerebellum *7*, 433–443.

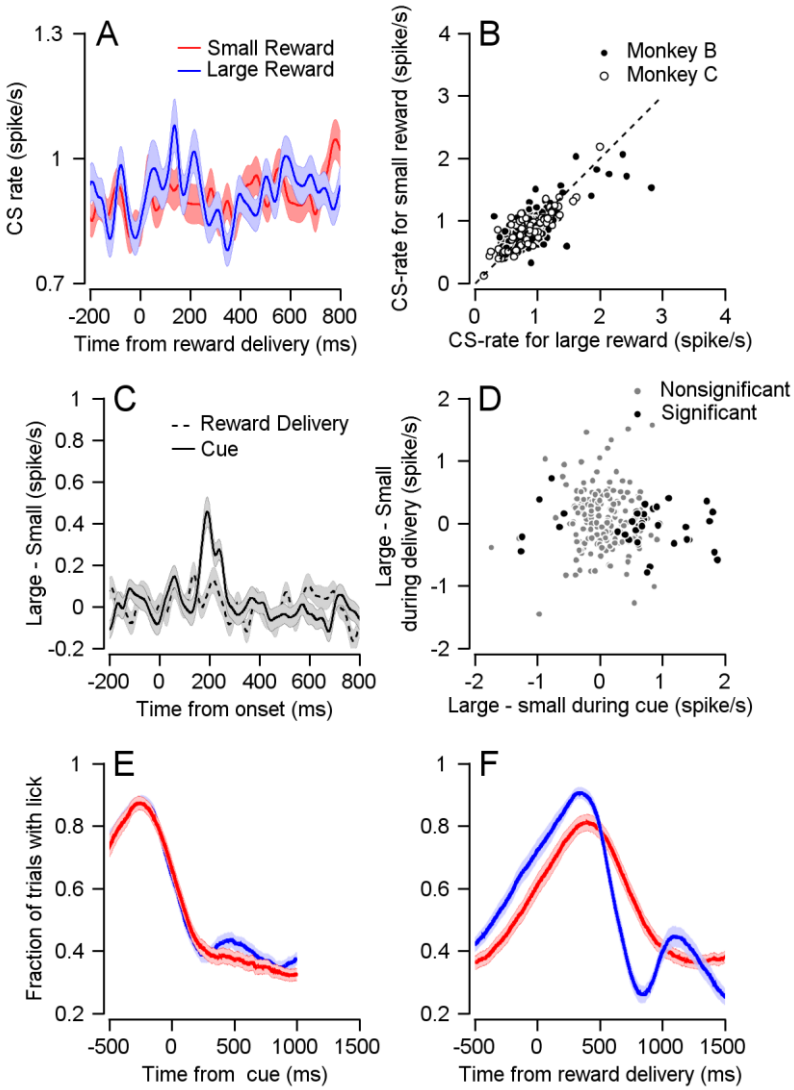
Thorndike, E.L. (1898). Animal intelligence: An experimental study of the associative processes in animals. Psychol. Rev.

Wagner, M.J., Kim, T.H., Savall, J., Schnitzer, M.J., and Luo, L. (2017). Cerebellar granule cells encode the expectation of reward. Nature *544*, 96–100.

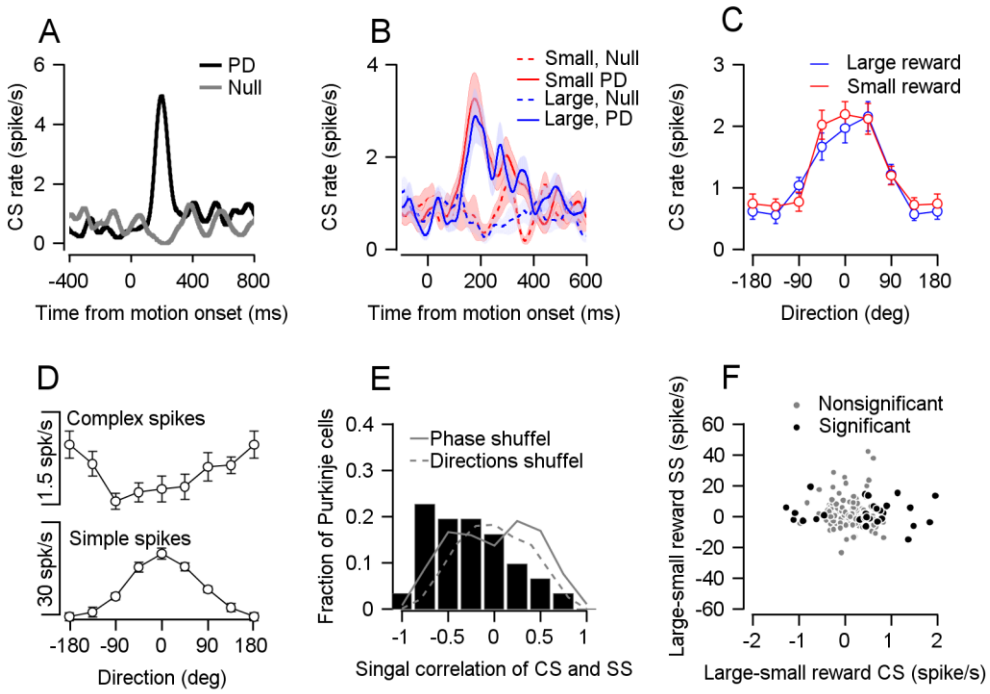
Welsh, J.P., Lang, E.J., Suglhara, I., and Llinás, R. (1995). Dynamic organization of motor control within the olivocerebellar system. Nature *374*, 453–457.



<u>Figure 1:</u> CS rate differentiates reward conditions during cue presentation. A, Eye movement task temporally separates reward expectation, pursuit behavior and reward delivery. **B** Traces of average eye speed, in the first 300ms after target motion onset. Target velocity was 20 °/s. **C**, Each dot represents the average speed for an individual session 250ms after target movement onset in the large (horizontal) and small (vertical) reward cue (Signrank, P=2*10^-18, n=115). **D**, Example cell PSTH. **E**, Population PSTH, CS rate increases in the large reward condition. In all figures the error bars represent SEM. **F**, Each dot represents the average CS rate of an individual cell 100-300 ms after the display of the large (horizontal) and small (vertical) reward cue (Signrank, Monkey B: P=0.01, n=148, Monkey C: P=7*10^-5, n=72).



<u>Figure 2:</u> CS firing pattern is not related to motor response to reward. A, Population PSTHs for different reward conditions overlap when aligned them to reward delivery. **B**, Each dot represents the average CS rate of an individual cell 100-300 ms large (horizontal) and small (vertical) reward delivery (Signrank, Monkey B: P=0.339, n=148; Monkey C: P=0.719, n=72). **C**, The differences between the PSTH for large and small rewards aligned to cue or to reward delivery. **D**, Each dot represents the average CS rate of an individual cell 100-300 ms after the cue (horizontal) and reward delivery (vertical). Spearman correlation of all cells: r=-0.0705, P=0.297, n=220; Spearman correlation of cells that responded to reward size during cue: r=-0.0805, P=0.62, n=40. **E** and **F**, Fraction of trials with licks, during cue and reward delivery.



<u>Figure 3:</u> Reward did not modulate CS direction tuning. **A**, PSTH of an example cell in its preferred and null directions **B**, Population PSTH for different reward conditions, in the preferred and null directions. **C**, Population direction tuning curve (Permutation test: P=0.2156, n=33). **D**, Population tuning curve of complex spikes (up) and simple spikes (bottom), both aligned to the preferred direction of simple spikes (Spearman r=-0.3087, P=7*10^-7, n=31). **E**, Histogram of signal correlations of simple and complex spikes in the population. Solid and dashed lines show the correlations on phased and direction shuffled data (Signrank: P= 0.002, n=31). **F**, Each dot shows individual cell differences in average rate 100-300 ms after cue, in CS (horizontal) and simple spikes (SS, vertical). Spearman correlation of all cells r=-0.0867, P=0.2580, n=172; Spearman correlation of cells that responded to reward size during cue: r=0.0323, P=0.8632, n=31.