Constructing threat probability, fear behaviour, and aversive prediction error in the brainstem

Authors: Jasmin A. Strickland[1]* and Michael A. McDannald[1]*

Affiliations:

[1]Boston College, Department of Psychology & Neuroscience, Boston College, Chestnut Hill, MA, 02467, USA

*Corresponding authors. Email: jasmin.strickland@bc.edu and michael.mcdannald@bc.edu

When faced with potential threat we must estimate its probability, respond advantageously, and leverage experience to update future estimates. Threat estimates are the proposed domain of the forebrain, while behaviour is elicited by the brainstem. Yet, the brainstem is also a source of prediction error, a learning signal to acquire and update threat estimates. Neuropixels probes allowed us to record single-unit activity across a 21-region brainstem axis during probabilistic fear discrimination. Against a backdrop of widespread threat probability and behaviour signaling, a dorsally-based brainstem network rapidly signaled threat probability. Remapping of neuronal function following shock outcome gave rise to brainstem networks signaling prediction error on multiple timescales. The results reveal construction of threat probability, behaviour, and prediction error along a single brainstem axis.

Introduction

Faced with potential threat, we must estimate its probability, determine an appropriate response, and – should we come away intact – adjust our estimates for future encounters. Historical and current descriptions of the brain's threat circuitry emphasize a division of labour in which forebrain regions estimate threat probability, while the brainstem elicits behaviour (*1*, *2*). Yet, behaviour signaling is not robustly observed in expected brainstem neuronal populations, such as the periaqueductal gray (*3*, *4*), which instead signals threat probability (*5*). Further, the brainstem periaqueductal gray is a source of prediction error (*3*, *6*, *7*), a learning signal to adjust threat estimates (*8*). These findings necessitate a more complex role for the brainstem in threat. However, evidence of widespread brainstem threat probability signaling remains elusive, and complete descriptions of brainstem behaviour and prediction error signaling are absent. Recording a 21-region axis with Neuropixels (*9*) during probabilistic fear discrimination (*10*), we report the brainstem constructs complete signals for threat probability, behaviour, and prediction error from neuronal 'building blocks' organized into functional networks. Remapping of neuron function between cue and post-shock periods revealed distinct brainstem network organization for threat probability and prediction error.

Results

Rats received probabilistic fear discrimination in which three cues predicted unique foot shock probabilities: danger (1), uncertainty (0.25) and safety (0) Fig. 1A). Rats were then implanted with a Neuropixels probe through the brainstem (Fig. 1B) to permit high-density, single-unit recordings from a complete dorsal-ventral axis during discrimination. Fear was measured via suppression of reward seeking, providing an index of rat's total behavioural knowledge of the cue-shock relationships. Rats showed complete, differential fear that was high to danger, intermediate to uncertainty, and low to safety (Fig. 1C; ANOVA main effect of cue, $F_{(2,142)} = 149.2$, $p=1.26 \times 10^{-35}$; Fig S1). We isolated and held 1,812 neurons from 10 rats during 75, 1-hr recording sessions (965 neurons from 4 females, Table S1). Neurons spanned 21 brainstem regions (Fig. 1D), including subregions and neighbouring regions of the superior colliculus, periaqueductal gray, dorsal raphe, and median raphe (Fig. 1E, Fig S2, Table S2).

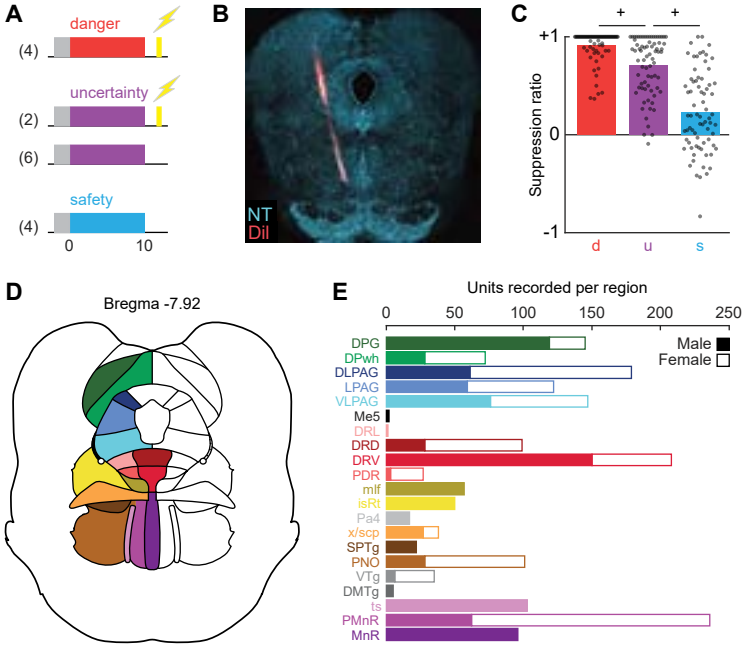

Figure 1. Fear discrimination and Neuropixels implant. (**A**) Probabilistic fear discrimination procedure. (**B**) Representative Neuropixels implant. (**C**) Cue suppression ratios during recording sessions. (**D**) Summary of brainstem regions recorded. (**E**) Number of single units recorded from each brainstem region by sex. $^{+}$95% bootstrap confidence interval does not contain zero.

Neurons obtained from each brainstem region showed marked cue firing that varied in time course,

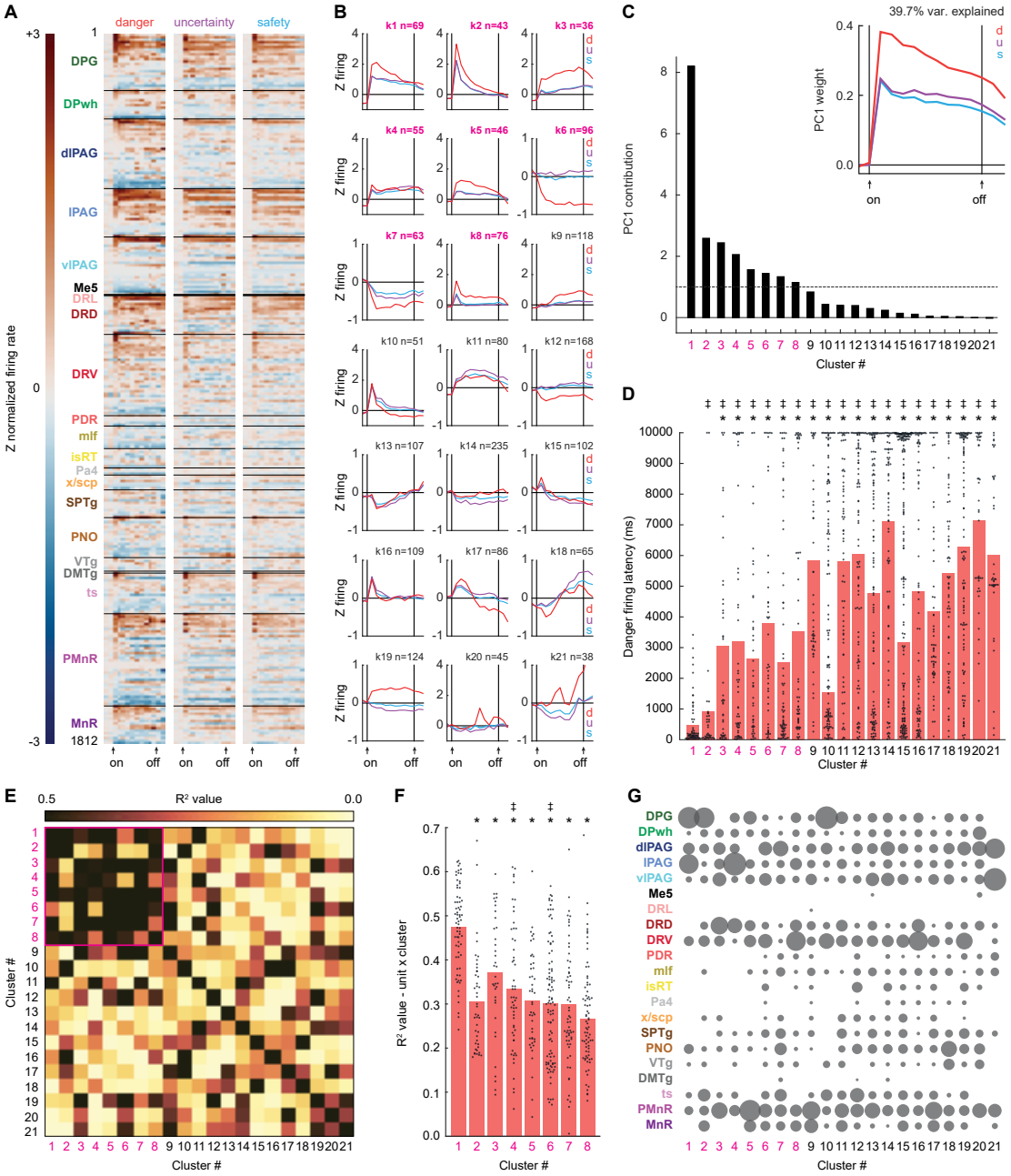

Figure 2. Brainstem cue firing. (**A**) Single-unit firing to danger, uncertainty, and safety cues organized by brain region, dorsal to ventral. (**B**) Mean cluster (k1-k21) firing over cue presentation. (**C**) PC1 for brainstem cue firing (inset) and cluster contribution to PC1. (**D**) Single-unit danger firing latency with cluster means. (**E**) Between-cluster cue firing correlations (data from B). Cue subnetwork outlined in magenta. (**F**) Single unit x cluster firing correlations for the cue subnetwork, with cluster means. (**G**) Proportion of each cluster found in each brainstem region. *Significance of independent samples Bonferroni-corrected t-test. ‡Significance of Levene's test for equality of variance, Bonferroni corrected.

direction, and specific cue pattern (Fig. 2A). K-means clustering revealed neurons could be organized into at least 21 functional clusters; potential building blocks for brainstem construction of threat probability and behaviour. Cluster size varied modestly (min size = 36, max = 235, and median = 76) and consistent firing themes emerged when clusters were visualized (Fig. 2B, Fig S3). Many clusters showed ordered cue firing that strongly differentiated danger and uncertainty, but modestly differenti-

ated uncertainty and safety. Principal component analysis (PCA) revealed ordered cue firing (danger > uncertainty > safety) to be the primary low-dimensional firing feature across all brainstem neurons (PC1, explaining 39.7% of firing variance; Fig. 2C, inset).

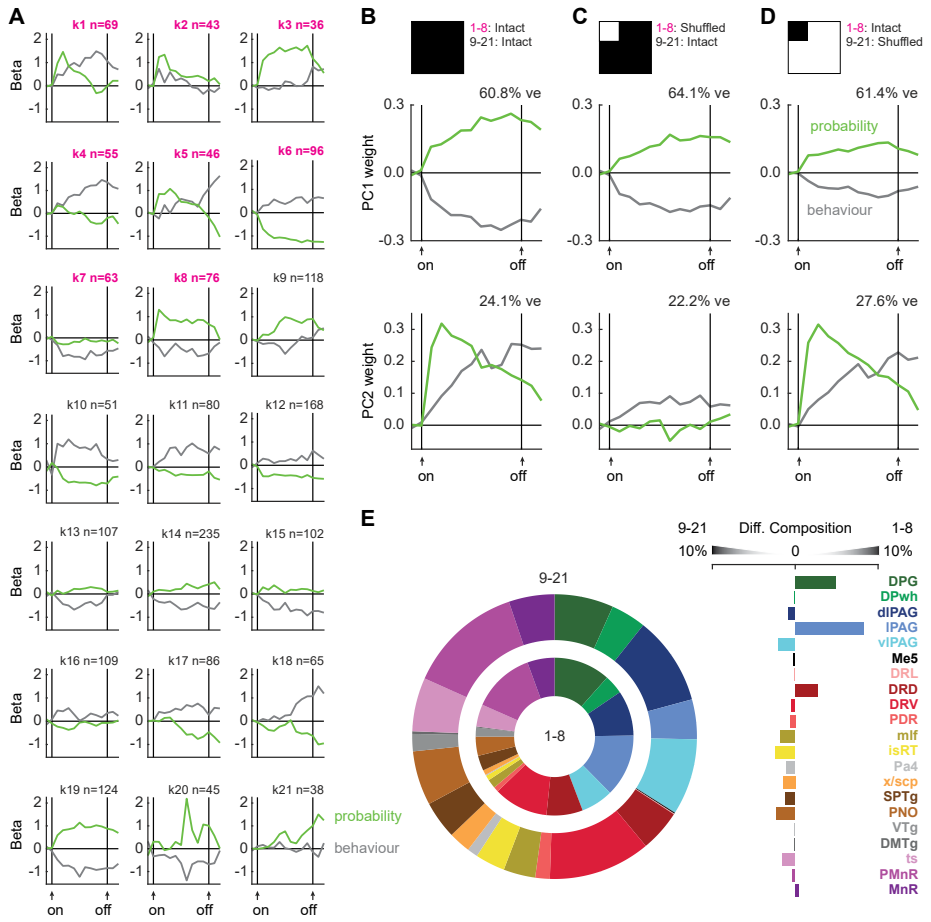

Figure 3. Brainstem threat and behaviour signaling. (**A**) Mean cluster (k1-k21) beta weights for threat probability and behaviour over cue presentation. (**B**) Principal components for cluster beta weights (top; data from A), resulting PC1 (middle), and PC2 (bottom). (**C**) Principal components for cluster beta weights with cue subnetwork shuffled (top; data from A), resulting PC1 (middle), and PC2 (bottom). (**D**) Principal components for cluster beta weights with cue supranetwork shuffled (top; data from A), resulting PC1 (middle), and PC2 (bottom). (**E**) Proportion of cue subnetwork and supranetwork single units by brain region (left), and differential composition of subnetwork and supranetwork by brain region (right).

But did all neuron types equally contribute to PC1 firing information? We utilized an iterative, PCA shuffle analysis to determine each cluster's PC1 contribution. Cue firing for the neurons comprising each specific cluster (e.g., k1) was shuffled, while cue firing for the neurons of all other clusters was left intact (e.g., k2-k21). Shuffling and PCA were performed 1000 times per cluster. The change in % explained firing variance from the complete data (39.7%) to the shuffled data was calculated and averaged across the 1000 iterations for that specific cluster [PC1 complete – mean (PC1 K1 shuffled)]. Clusters contributing more greatly to PC1 have higher values. PC1 firing information largely originated from eight clusters (k1-k8; Fig. 2C, Fig S4) composed of a minority of neurons (484/1812, 26.7%). Clusters k1-k8 showed shorter firing latencies to danger onset, which correlated with their PC1 contribution ($R^2 = 0.44$, p=0.001; Fig. 2D). These eight clusters further separated themselves based on intra-cluster correlation of cue firing (Fig. 2E. Fig S4), forming a functional subnetwork within the larger brainstem network. K1 neurons showed indicators of a subnetwork hub: having the greatest PC1 contribution, least variation and the shortest mean danger firing latency (Fig. 2D). Further, k1 single-unit firing correlated most strongly with population firing of their fellow subnetwork clusters (Fig. 2F, Fig S4).

Neurons from each cluster were observed in at least eight brainstem regions (Fig. 2G). Subnetwork neurons were concentrated in four, core regions: the deep layer of the superior colliculus, lateral sub-

division of the periaqueductal gray, dorsal subdivision of the dorsal raphe, and paramedian raphe. Subnetwork hub neurons (k1) were particularly concentrated in the deep layer of the superior colliculus and the lateral subdivision of the periaqueductal gray.

Ordered cue firing is the predominant brainstem activity feature. Yet, ordered cue firing could equally reflect threat probability or behaviour. Cue firing reflecting threat probability should linearly scale with foot shock probability (0.0, 0.25, and 1.0), invariant of the level of cued fear. Conversely, cue firing reflecting behaviour should reflect the level of cued fear, invariant of foot shock probability. Linear regression revealed widespread, yet unique threat probability and behaviour signals across the 21 clusters (Fig. 3A). PCA for cluster beta coefficients (Fig. 3B, top) revealed PC1 to reflect sustained, opposing signals for threat probability and behaviour (60.9% of signaling variance; Fig. 3B, middle). PC2 reflected dynamic, probability-to-behaviour signaling (24.1% of signaling variance; Fig. 3B, bottom, Fig S4). To reveal network-specific contributions to threat probability and behaviour signaling, we iteratively shuffled or 'lesioned' cluster firing for one network (e.g., subnetwork clusters k1-k8), while leaving the remaining clusters intact (e.g., supranetwork clusters k9-k21). Linear regression was performed for each cluster, then PCA was performed for all clusters to reveal low-dimensional signaling features. Comparing fully intact signaling (Fig. 3B), signaling with the subnetwork 'lesioned' (Fig. 3C), versus the cue supranetwork 'lesioned' (Fig. 3D) allowed us to determine the relative contributions of each network to threat probability and behaviour signaling.

Sustained behaviour signaling depended more on the cue supranetwork, while dynamic, probability-to-behaviour signaling depended entirely on the cue subnetwork (Fig. 3C and D). Lesioning the subnetwork left sustained threat probability and behaviour signaling largely intact (Fig. 3C, middle), but abolished dynamic, probability-to-behaviour signaling (Fig. 3C, bottom). By contrast, lesioning the supranetwork most greatly diminished sustained behaviour signaling (Fig. 3D, middle), but left dynamic, probability-to-behaviour signaling fully intact (Fig. 3D, bottom).

Brainstem regions differed in their percent composition of subnetwork vs. supranetwork cluster neurons (Fig. 3E). The deep layer of the superior colliculus, lateral subdivision of the periaqueductal gray, and dorsal subdivision of the dorsal raphe preferentially contributed to the subnetwork. Ventral brainstem regions preferentially yet more modestly contributed to the supranetwork. These findings reveal the brainstem constructs complete signals for threat probability and behaviour through partially distinct functional networks. A more dorsal brainstem network constructs a rapid threat probability signal from a subset of neuronal building blocks. A diffuse, more ventral brainstem network constructs continuous threat probability and behaviour signals from separate neuronal building blocks.

Threat probability information contained in cue firing is shaped by prediction error – a learning signal generated following shock delivery and omission. To capture prediction error-related firing, we focused on the 10 s following shock offset. Brainstem neurons showed marked yet varied firing changes, particularly following 'surprising' foot shock on uncertainty trials (Fig. 4A). K-means clustering revealed brainstem neurons could be organized into at least 11 functional clusters (min size = 27, max = 356, and median = 76; Fig. 4B, Fig S5). Strikingly, PCA revealed signed prediction error to be the primary low-dimensional firing feature across all brainstem neurons (PC1 explains 20.2% of firing variance,

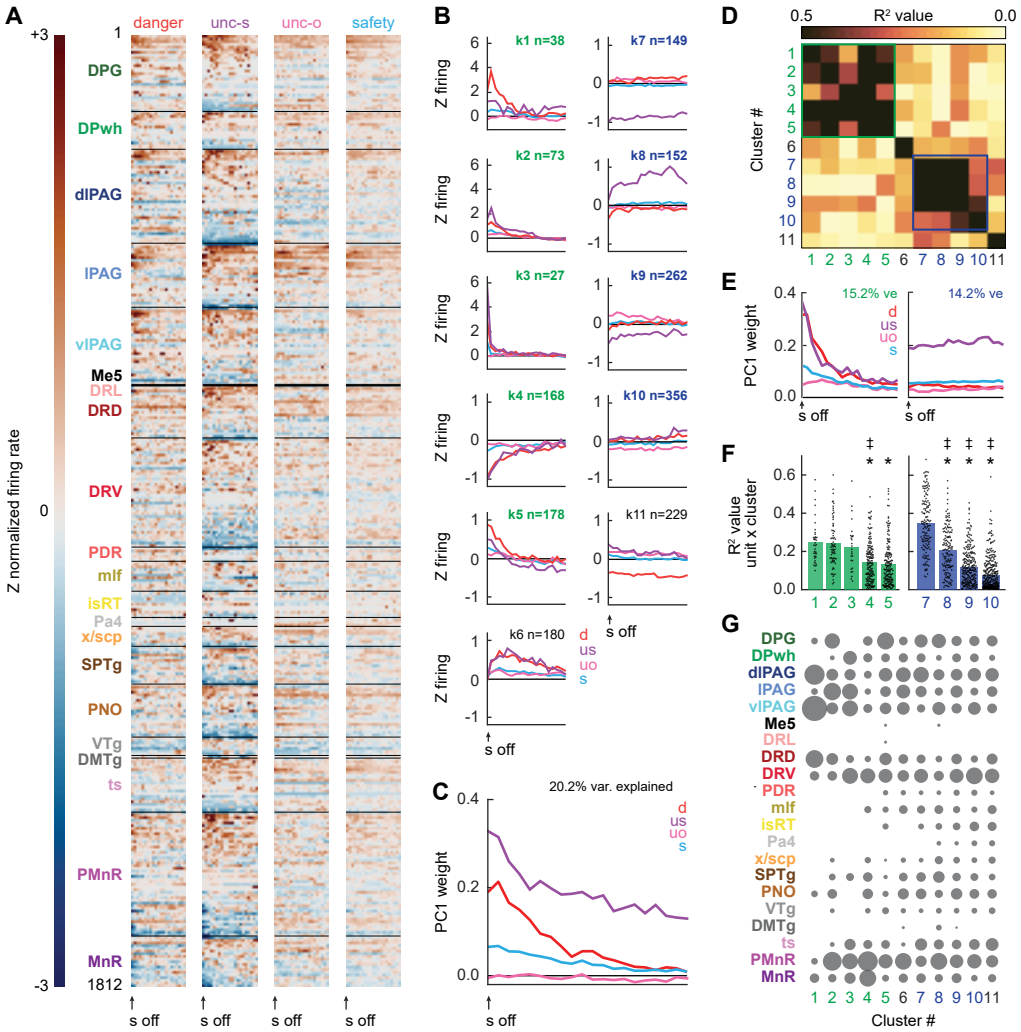

Figure 4. Brainstem outcome firing. (**A**) Single-unit firing following danger, uncertainty shock, uncertainty omission, and safety organized by brain region, dorsal to ventral. (**B**) Mean cluster (k1-k11) firing following foot shock. (**C**) PC1 for brainstem outcome firing. (**D**) Between-cluster cue firing correlations (data from B). Phasic outcome network outlined in green, tonic outcome network outlined in blue. (**E**) PC1 for phasic outcome network firing (left), and tonic outcome network firing (right). (**F**) Single unit x cluster firing correlations for the phasic outcome network (left, green), and tonic outcome network (right, blue). (**G**) Proportion of each cluster found in each brainstem region. *Significance of independent samples Bonferroni-corrected t-test. ‡Significance of Levene's test for equality of variance, Bonferroni corrected.

Fig. 4C, Fig S6). The prediction error is 'signed' because the PC1 weight for surprising shock exceeds that for predicted shock (following danger) – the positive component of signed error. Concurrently, the PC1 weight for surprising omission is lesser than that for predicted omission (following safety) – the negative component of signed error. Shock responding was also evident, as the PC1 weight for predicted shock exceeded that for predicted omission.

How does the brainstem construct prediction error from underlying neuronal building blocks? Unlike the cue period, the temporal profile of shock firing was the organizing principle for outcome clusters. Neurons composing clusters k1-k5 showed transient firing changes following foot shock, although k1 & k5 neurons preferentially fired to predicted shock, while k2 & k3 neurons preferentially fired to surprising shock (Fig. 4B, left column). Highly correlated temporal firing revealed a phasic outcome network

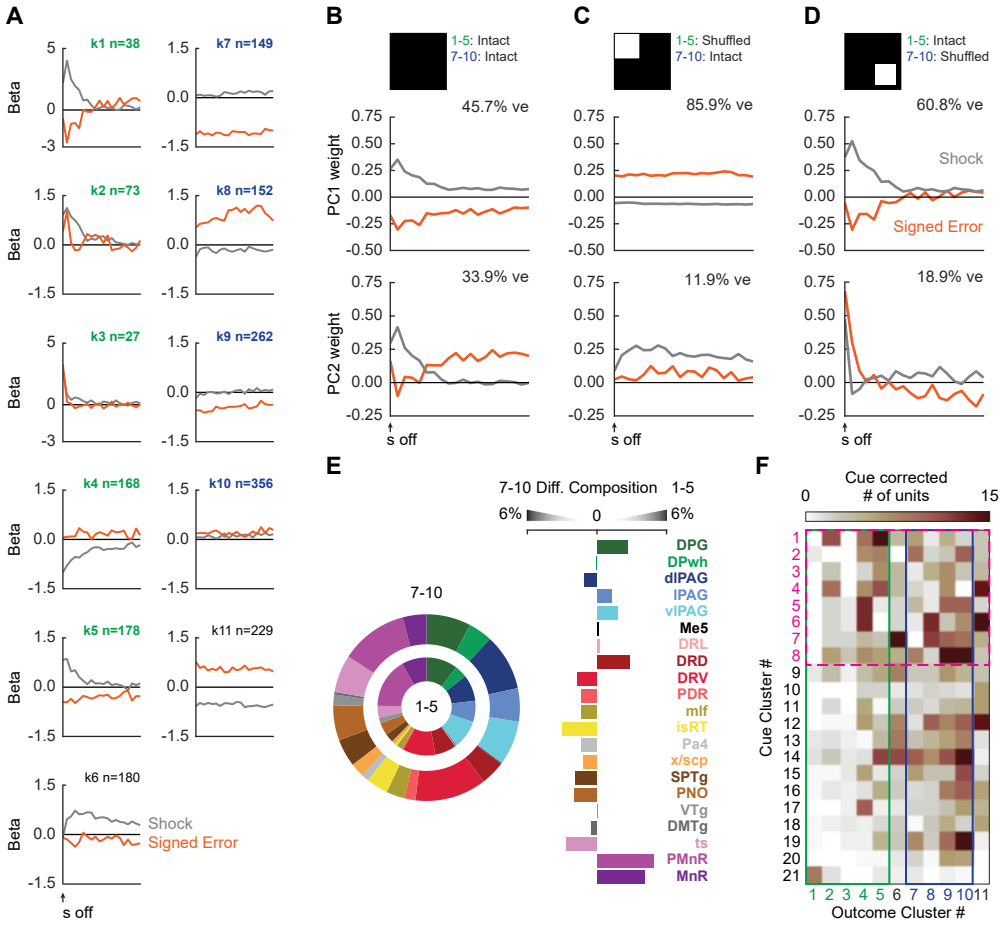

Figure 5. Brainstem shock and prediction error signaling. (**A**) Mean cluster (k1-k11) beta weights for shock and prediction error following shock delivery and omission. (**B**) Principal components for cluster beta weights (top; data from A), resulting PC1 (middle), and PC2 (bottom). (**C**) Principal components for cluster beta weights with phasic outcome network shuffled (top; data from A), resulting PC1 (middle), and PC2 (bottom). (**D**) Principal components for cluster beta weights with tonic outcome network shuffled (top; data from A), resulting PC1 (middle), and PC2 (bottom). (**E**) Proportion of phasic and tonic network single units by brain region (left), and differential composition of tonic and phasic network by brain region (right). (**F**) Relationship between cue cluster and outcome cluster membership across all brainstem neurons.

(Fig. 4D, Fig S6). Neurons composing clusters k7-k10 showed sustained firing changes following foot shock (Fig. 4B, right column), with highly correlated firing revealing a tonic outcome network (Fig. 4D). The two outcome networks emphasized different features of prediction error. PC1 for the phasic outcome network was transient foot shock responding, rather than positive prediction error; combined with differential responding to predicted and surprising omission, negative prediction error (explaining 15.2% of firing variance; Fig. 4E, left). PC1 for the tonic outcome network emphasized positive error, with negative error apparent but diminished (explaining 14.2% of firing variance; Fig. 4E, right).

The phasic outcome network lacked a clear hub. Clusters k1, k2 and k3 – candidate hubs – were each composed of neurons whose firing correlated equally well with population firing of their fellow network clusters (Fig. 4F, left). By contrast, k7 neurons were a hub for the tonic outcome network. Firing of k7 neurons, the cluster strongly decreasing firing to surprising shock, correlated most strongly with population firing of their fellow tonic outcome clusters (Fig. 4F, right). Neuronal populations that differed in

their temporal profile differed in their anatomical distribution (Fig. 4G). Phasic outcome neurons were more common at axis extremes: subregions of the periaqueductal gray, dorsal raphe, and median raphe. By contrast, tonic outcome neurons were more evenly distributed across the brainstem. This distribution was most striking for tonic outcome hub neurons which were distributed nearly evenly across the brainstem axis.

Yet, how is foot shock versus prediction error information organized in each network? Linear regression revealed unique foot shock and prediction error signals across the 11 outcome clusters (Fig. 5A). PCA for cluster beta coefficients revealed opposing, phasic then tonic shock and prediction error to be the primary low-dimensional signaling feature across all brainstem neurons (PC1, 45.7% of signaling variance; Fig. 5B, middle). PC2 reflected dynamic, shock-to-signed prediction error signaling (33.9% of signaling variance; Fig. 5B, bottom, Fig S6). We again turned to PCA lesion analysis to reveal the relative contributions of the phasic and tonic outcome networks to signaling information. Rapid shock signaling depended on the phasic outcome network (Fig. 5C). Lesioning the phasic outcome network left signaling dominated by sustained prediction error (85.9% of signaling variance; Fig. 5C, middle), while residual signaling reflected sustained shock (Fig. 5C, bottom). By contrast, lesioning the tonic outcome network emphasized opposing, phasic shock and prediction error signaling (60.8% of signaling variance; Fig. 5D, middle). Now, residual signaling reflected phasic, unidirectional shock and prediction error (Fig. 5D, bottom). Thus, a phasic brainstem network constructs signals for shock *and* prediction error from a subset neuronal building blocks transiently active following shock. Concurrently, a tonic brainstem network preferentially constructs prediction error from a subset of neuronal building blocks sustaining activity following shock.

Anatomical biases for outcome network neurons were less striking than those for cue subnetwork and supranetwork. The phasic outcome network was situated at the dorsal and ventral extremes: deep layer of the superior colliculus, lateral and ventrolateral periaqueductal gray, dorsal subdivision of the dorsal raphe, and paramedian/median raphe. The tonic outcome network resided in central brainstem regions situated between the dorsal and median raphe. Finally, it is intriguing that the same 1,812 neurons constructing threat probability and behaviour during cue presentation, constructed shock and prediction error following outcome presentation. We were curious whether there was a relationship between network membership during cue and outcome periods. The cue subnetwork was composed of 484 neurons (484/1812, 27.6%). Phasic outcome neurons were more likely to be observed in the cue subnetwork (169/484, 34.9%; $\chi^2 = 16.19$, $p<0.0001$), while tonic outcome neurons were less likely to be observed in the cue subnetwork (190/919, 20.7%; $\chi^2 = 16.84$, $p<0.0001$). The brainstem cue subnetwork constructing threat probability is most distinct from the tonic outcome network constructing prediction error.

Discussion

We set out to reveal brainstem construction of threat probability versus behaviour. Supporting a prevailing view (*1*), we observed brainstem functional populations whose firing was better captured by trial-by-trial fluctuations in behaviour, rather than threat probability. The firing independence of these populations and their diffuse anatomical distribution meant continuous brainstem behaviour signaling from cue onset until shock delivery. Opposing the prevailing view, brainstem populations whose firing was better captured by threat probability were equally numerous. Continuous brainstem threat probability signaling was also achieved by anatomically diffuse functional populations. However, a highly organized threat probability signal was also uncovered. Brainstem populations showing pronounced differential firing to danger and safety, plus short-latency danger firing changes, formed a functional network. Neurons contributing to this network were focused in dorsal brainstem regions, with 'hub' neurons concentrated in the deep layer of the superior colliculus and the lateral periaqueductal gray. Most novel, network firing at cue onset preferentially signaled threat probability, giving way to behaviour signaling as foot shock drew near. Rather than being the exclusive domain of the forebrain, the brainstem constructs, and even prioritizes, threat probability.

We further found that prediction error signaling is fundamental to the brainstem. This is broadly consistent with prior studies which have reported prediction error in the periaqueductal gray (*6*, *7*, *11*). However, our results reveal a more complex picture. First, the brainstem contains two prediction error networks operating on different time scales. A phasic outcome network is rapidly engaged following shock, with composing functional populations generally responsive to shock, or showing selective responding to surprising or predicted shock. Populations for surprising shock likely correspond to known centers for prediction error generation (*7*). A tonic outcome network specifically signals prediction error. Unique to the tonic outcome network: composing functional populations – even the hub – are highly anatomically distributed. Even more, hub neurons show preferential firing *decreases* to surprising shock. The results suggest there may be multiple brainstem prediction error systems. Alternatively, there may be a single prediction error system in which a tonic signal opens a window of permissibility for phasic prediction error to update threat estimates (*12*).

Viewing the forebrain as *the* source of threat estimation has meant continuous refinement of forebrain threat processing. Cortical subregions are linked to increasingly specific threat functions (*13*). Amygdala threat microcircuits are being mapped in intricate detail (*14*). Brainstem regions contain the building blocks needed to construct threat estimates. This finding necessitates refinement and detail of brainstem threat function on par with its forebrain counterparts. Expanding on prior brainstem work (*15*–*17*), our results reveal the superior colliculus (*18*–*20*) and periaqueductal gray (*21*–*23*) as prominent sources of threat information. Somewhat unexpectedly, we reveal abundant and diverse threat signaling in the paramedian raphe (*24*), a virtually unstudied region adjacent to the serotonin-containing median raphe. Most critically, these regions do not function in isolation. Rather, the superior colliculus and periaqueductal gray organize a local brainstem network to rapidly signal threat probability.

Perhaps the brainstem signals threat probability, but this signal is trained up by the forebrain. This would be consistent with our findings. Yet, where do forebrain threat estimates come from? Once

formed, how are threat estimates updated? Prediction error provides a plausible mechanism for forming and updating threat estimates. Preferential responding to surprising aversive events – consistent with positive prediction error – has been reported in many forebrain regions (*25*). Preferential responding to omission of aversive events – consistent with negative error – has also been observed (*26*). However, opposing firing changes to positive and negative error in the same neuronal population – a requirement of a *fully signed* prediction error (*27*, *28*) – are more narrowly observed in the brainstem. Learned threat estimates originating in the forebrain then require prediction error generated in the brainstem. In which case, *de novo* acquisition of a brainstem threat estimate, trained by the forebrain, would require a brainstem-generated prediction error. Similarly, brainstem prediction error may be necessary to update forebrain threat estimates in the face of changing threat contingencies. Equally plausible – brainstem-generated prediction error may train and update a brainstem threat estimate, bypassing the forebrain altogether.

Fully revealing the brain basis of threat computation is essential to understanding healthy and disordered fear. Our finding of widespread and organized brainstem threat signaling calls for abandonment of the historical division of labour view. In its place we must embrace a brain-wide view of threat computation in which brainstem networks are integral to constructing threat.

Acknowledgments

We thank Bret Judson and the Boston College Imaging Core for infrastructure/support, Dr. Matthew Gardner and Dr. Geoffrey Schoenbaum for advice and initial designs for 3D head cap printing, Joe Austen for help building and ordering acquisition computers, Richard Pijar and the Boston College Machine Shop for electrical and hardware support and instrument customization, Pavel Kulik and Open Ephys for assistance with the I/O module, Dr. Nicholas Steinmetz, Dr. Matteo Carandini and the UCL Neuropixels course for initial training in the use of silicone probes. Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under Award Number R01MH117791. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors report no competing interests. This work was also supported in part by an Ignite Grant from Boston College.

Author contributions

Conceptualization: JS, MM
Methodology: JS, MM
Investigation: JS, MM
Funding acquisition: MM
Writing – original draft: JS, MM
Writing – review and editing: JS, MM

Competing interests

Authors declare that they have no competing interests.

References

1. I. Levy, D. Schiller, Neural Computations of Threat. *Trends Cogn Sci*. **25**, 151–171 (2021).
2. M. S. Fanselow, Neural Organization of the Defensive Behavior System Responsible for Fear. *Psychon B Rev*. **1**, 429–438 (1994).
3. T. Ozawa, E. A. Ycu, A. Kumar, L. F. Yeh, T. Ahmed, J. Koivumaa, J. P. Johansen, A feedback neural circuit for calibrating aversive memory strength. *Nat Neurosci*. **20**, 90–97 (2017).
4. P. Tovote, M. S. Esposito, P. Botta, F. Chaudun, J. P. Fadok, M. Markovic, S. B. Wolff, C. Ramakrishnan, L. Fenno, K. Deisseroth, C. Herry, S. Arber, A. Luthi, Midbrain circuits for defensive behaviour. *Nature*. **534**, 206–12 (2016).
5. K. M. Wright, M. A. McDannald, Ventrolateral periaqueductal gray neurons prioritize threat probability over fear output. *Elife*. **8** (2019), doi:10.7554/eLife.45013.
6. R. A. Walker, K. M. Wright, T. C. Jhou, M. A. McDannald, The ventrolateral periaqueductal gray updates fear via positive prediction error. *Eur J Neurosci* (2019), doi:10.1111/ejn.14536.
7. M. Roy, D. Shohamy, N. Daw, M. Jepma, G. E. Wimmer, T. D. Wager, Representation of aversive prediction errors in the human periaqueductal gray. *Nature neuroscience*. **17**, 1607–12 (2014).
8. R. A. Rescorla, A. R. Wagner, in *Classical Conditioning II: Current Research and Theory*, B. AH, P. WF, Eds. (Appleton Century Crofts, New York, 1972), pp. 64–99.
9. J. J. Jun, N. A. Steinmetz, J. H. Siegle, D. J. Denman, M. Bauza, B. Barbarits, A. K. Lee, C. A. Anastassiou, A. Andrei, C. Aydin, M. Barbic, T. J. Blanche, V. Bonin, J. Couto, B. Dutta, S. L. Gratiy, D. A. Gutnisky, M. Hausser, B. Karsh, P. Ledochowitsch, C. M. Lopez, C. Mitelut, S. Musa, M. Okun, M. Pachitariu, J. Putzeys, P. D. Rich, C. Rossant, W. L. Sun, K. Svoboda, M. Carandini, K. D. Harris, C. Koch, J. O'Keefe, T. D. Harris, Fully integrated silicon probes for high-density recording of neural activity. *Nature*. **551**, 232–236 (2017).
10. M. Moaddab, M. A. McDannald, Retrorubral field is a hub for diverse threat and aversive outcome signals. *Current Biology*. **31**, 2099-2110.e5 (2021).
11. J. P. Johansen, J. W. Tarpley, J. E. LeDoux, H. T. Blair, Neural substrates for expectation-modulated fear learning in the amygdala and periaqueductal gray. *Nature neuroscience*. **13**, 979–86 (2010).
12. A. A. Grace, Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: A hypothesis for the etiology of schizophrenia. *Neuroscience*. **41**, 1–24 (1991).
13. M. Alexandra Kredlow, R. J. Fenster, E. S. Laurent, K. J. Ressler, E. A. Phelps, Prefrontal cortex, amygdala, and threat processing: implications for PTSD. *Neuropsychopharmacol.*, 1–13 (2021).
14. S. Duvarci, D. Pare, Amygdala microcircuits controlling learned fear. *Neuron*. **82**, 966–80 (2014).
15. K. Kveraga, J. Boshyan, R. B. Adams, J. Mote, N. Betz, N. Ward, N. Hadjikhani, M. Bar, L. F. Barrett, If it bleeds, it leads: separating threat from mere negativity. *Soc Cogn Affect Neur*. **10**, 28–35 (2015).
16. B. J. Liddell, K. J. Brown, A. H. Kemp, M. J. Barton, P. Das, A. Peduto, E. Gordon, L. M. Williams, A direct brainstem-amygdala-cortical "alarm" system for subliminal signals of fear. *Neuroimage*. **24**, 235–243 (2005).
17. J. M. P. Baas, J. Milstein, M. Donlevy, C. Grillon, Brainstem Correlates of Defensive States in Humans. *Biological Psychiatry*. **59**, 588–593 (2006).
18. P. A. Kragel, M. Čeko, J. Theriault, D. Chen, A. B. Satpute, L. W. Wald, M. A. Lindquist, L. Feld-

man Barrett, T. D. Wager, A human colliculus-pulvinar-amygdala pathway encodes negative emotion. *Neuron*. **109**, 2404-2412.e5 (2021).

19. Y. C. Wang, M. Bianciardi, L. Chanes, A. B. Satpute, Ultra High Field fMRI of Human Superior Colliculi Activity during Affective Visual Processing. *Sci Rep*. **10**, 1331 (2020).
20. Z. Zhao, M. Davis, Fear-Potentiated Startle in Rats Is Mediated by Neurons in the Deep Layers of the Superior Colliculus/Deep Mesencephalic Nucleus of the Rostral Midbrain through the Glutamate Non-NMDA Receptors. *J. Neurosci.* **24**, 10326–10334 (2004).
21. O. K. Faull, M. Jenkinson, M. Ezra, Kt. Pattinson, Conditioned respiratory threat in the subdivisions of the human periaqueductal gray. *Elife*. **5** (2016), doi:10.7554/eLife.12047.
22. A. B. Satpute, T. D. Wager, J. Cohen-Adad, M. Bianciardi, J. K. Choi, J. T. Buhle, L. L. Wald, L. F. Barrett, Identification of discrete functional subregions of the human periaqueductal gray. *P Natl Acad Sci USA*. **110**, 17101–17106 (2013).
23. G. P. McNally, S. Cole, Opioid receptors in the midbrain periaqueductal gray regulate prediction errors during pavlovian fear conditioning. *Behavioral neuroscience*. **120**, 313–23 (2006).
24. K. E. Sos, M. I. Mayer, C. Cserép, F. S. Takács, A. Szőnyi, T. F. Freund, G. Nyiri, Cellular architecture and transmitter phenotypes of neurons of the mouse median raphe region. *Brain Struct Funct*. **222**, 287–299 (2017).
25. A. Ploghaus, I. Tracey, S. Clare, J. S. Gati, J. N. P. Rawlins, P. M. Matthews, Learning about pain: The neural substrate of the prediction error for aversive events. *P Natl Acad Sci USA*. **97**, 9281–9286 (2000).
26. V. I. Spoormaker, K. C. Andrade, M. S. Schroter, A. Sturm, R. Goya-Maldonado, P. G. Samann, M. Czisch, The neural correlates of negative prediction error signaling in human fear conditioning. *Neuroimage*. **54**, 2250–2256 (2011).
27. W. Schultz, P. Dayan, P. R. Montague, A neural substrate of prediction and reward. *Science*. **275**, 1593–9 (1997).
28. M. R. Roesch, D. J. Calu, G. Schoenbaum, Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*. **10**, 1615–24 (2007).
29. M. Pachitariu, N. Steinmetz, S. Kadir, M. Carandini, K. Harris, Fast and accurate spike sorting of high-channel count probes with KiloSort. *Adv Neur In*. **29** (2016) (available at ://WOS:000458973702073).
30. F. Crameri, Scientific colour maps (Version 4.0.0). (2018), , doi:10.5281/zenodo.2649252.

Materials and Methods

Subjects

Subjects were six male and four female Long-Evans rats, split over two rounds of testing. The first round included three female and two male rats born in the Boston College Animal Care facility, housed with mothers until postnatal day 21 when they were weaned and single housed. The second round included four males and one female, obtained from Charles River weighing 250g-275g on arrival. All were maintained on a 12-hour light-dark cycle (lights on 0600–1800) and were aged between 95 - 140 days old at the time of first recording session. All protocols were approved by the Boston College Animal Care and Use Committee, and all experiments were carried out in accordance with the NIH guidelines regarding the care and use of rats for experimental procedures.

Behavioural apparatus

Training took place in individual sound-attenuated enclosures that each housed a behaviour chamber with aluminum front and back walls, clear acrylic sides and top, and a metal grid floor. Each grid floor bar was electrically connected to an aversive shock generator (Med Associates, St. Albans, VT) through a device that ensured the floor was always grounded apart from during shock delivery. A single food cup and central nose poke opening equipped with infrared photocells were present on one wall. Auditory stimuli were presented through two speakers mounted on the enclosure ceiling. Auditory cues were 10s in duration and consisted of repeating motifs of a broadband click, phaser, or trumpet, which previous studies have found to be discriminable and equally salient. Testing took place in an identical chamber, but was equipped with a custom plastic food cup, plastic front and back walls, and multi-axis counterbalanced lever arm (Instech Laboratories, MCLA) with plastic tubing that held the recording cable and entered the chamber via a custom plastic top.

Nose poke acquisition

Rats were food restricted to 85% of their free-feeding body weight, with ad-libitum access to water. After pre-exposure to pellets (Bio-Serv, Flemington, NJ) in their home cages for two days, rats were shaped to nose poke for pellet in the experimental chamber. During the first session, the nose poke port was removed, and rats were issued one pellet every 60 seconds for 30 minutes. In the next session, the port was reinserted, and poking was reinforced on a fixed ratio 1 schedule in which one nose poke yielded one pellet until they reached ~50 nose pokes or 30min. Nose poking was then reinforced on a variable interval 30-second (VI-30) schedule for one session, then a VI-60 schedule for the next four sessions. The VI-60 reinforcement schedule was utilized during subsequent fear discrimination and was independent of auditory cue and foot shock presentation.

Fear discrimination

Rats received twelve sessions of Pavlovian fear discrimination prior to Neuropixels implant. Each 54-min session consisted of a five-minute warm up period in the chamber followed by 16 cue presentation trials. Each auditory cue predicted a unique foot shock probability (0.5 mA, 0.5 s): danger, p=1.00; uncertainty, p=0.25; and safety, p=0.00. Foot shock was administered two seconds following the termination of the cue on danger and uncertainty-shock trials. A single session consisted of 4 danger,

2 uncertainty-shock, 6 uncertainty-no shock, and 4 safety trials with a mean inter-trial interval of 3 min. Trial order was randomly determined by the behavioural program and differed for each rat, every session. The physical identities of the auditory cues were counterbalanced across individuals. Following recovery from surgery, rats received one VI-60 session to habituate to being connected to the recording cable. Rats then received between 1 and 10 discrimination sessions during which single-unit activity was recorded.

Surgery

Following the 12th discrimination session rats were returned to ad-libitum food access and underwent stereotaxic surgery performed under isoflurane anesthesia (1-5% in oxygen). Four screws were screwed into the skull around the target cap area to aid adhesion of the cap, and the skull was also scored in a crosshatch pattern. A craniotomy with a 1.4 mm diameter was carried out, and the underlying dura fully removed to expose the cortex. Immediately prior to implant the probe was painted with DiI to later identify histology tracks (ThermoFisher, V22886). To maximize recording regions, each implant was aimed at coordinates -8.00 AP, -2.80 ML, -7 to -7.5 DV, with a 15º angle. Each Neuropixels probe (1.0 probe) and head stage were secured in a pre-prepared custom head cap. The cap was held and slowly lowered during implant using a modified stereotaxic arm until the max DV was reached, or until the cap contacted the skull. The craniotomy was sealed using silicone gel (Dow DOWSIL 3-4680). Once the cap was in place, the ground wire was wrapped around the two screws positioned laterally to the cap to ground the probe. Vacuum sealing grease (Dow Corning) was applied around the base of the cap to fill any space between the cap and the skull and protect the probe. Caps were cemented into place using orthodontic resin (cc 22-05-98, Pearson Dental Supply) and the head cap lid secured in place on the head cap. Rats were given one week to recover with prophylactic antibiotic treatment (cephalexin, Henry Schein Medical) prior to data acquisition and received carprofen (5mg/kg) for post-operative analgesia.

Data acquisition

Neural data were recorded using OpenEphys with the Neuropixels PXI plugin running on an acquisition computer connected to the PXI chassis (PXIe-1071) containing the Neuropixels base station. Behaviour events were controlled and recorded by a separate computer running Med Associates software. To get behaviour timestamps, signals were sent from Med Associates to the NIDAQmx OpenEphys plugin, via Med Associates TTL adapter boxes (SG-231) plugged into a connector block (National Instruments, BNC 2110) connected to an I/O module (PXI-6363) in the PXI chassis. During recording sessions, the cable was first connected to the head stage and the head stage lid fixed in place, then the recording channels and reference for that session and subject were selected. To maximize acquisition of neurons from the midbrain region, the channels selected were either the lowest bank of 384 channels, or channels 193-575, used in a double alternating order across sessions and counterbalanced across subjects. The external reference was selected unless that proved ineffective in which case the tip reference of the probe was used instead. After this the doors to the chamber were closed and the fear discrimination and recording session started. Sessions were only included for analysis if the probe signal was maintained throughout all 16 trials, if the signal was lost for any reason that session was discarded. Subjects were recorded from daily up to either ten total recording sessions, or until data was

no longer able to be acquired from a subject.

Probe retrieval

Following recording sessions rats were placed back into the stereotaxic frame under isoflurane anesthesia. The head cap lid was removed, the ground wire cut, and the head stage disconnected and removed. The cement securing the probe holder in place was scraped away with a scalpel blade and the holder slowly pulled up and out of the cap. The probe was then rinsed and soaked in DI water, followed by a soak in a tergazyme solution before a final rinse with DI water before and if still functional after explant safely stored for re-implant.

Histology

Once the probe had been explanted, the rat was removed from the frame and deeply anesthetized using isoflurane before being perfused intracardially with 0.9% biological saline and 4% paraformaldehyde in a 0.2M potassium phosphate buffered solution. Brains were extracted and fixed in a 10% formalin solution for 24hrs, then stored in 10% sucrose/formalin. Brains were sliced with a microtome into forty micrometer sections (from approximately Bregma -6.5 to -9, to ensure the full extent of the probe tracks could be identified). The tissue was rinsed, incubated in NeuoroTrace (ThermoFisher, N21479), rinsed again, and then mounted prior to imaging within a week of processing (Axio Imager, Z2, Zeiss) to locate probe placement using the visible DiI tracks and NeuroTrace. Neuron locations were established by identifying the 3D location of the tip of the probe relative to the Allen Atlas, as well as the location in which the probe entered the brain (not including the cortex) and the vector of implant calculated. These were also checked against expected electrophysiology patterns (regions of expected low and high activity) for location accuracy.

Neuron sorting

See supplemental methods for full description. Data were automatically spike sorted using Kilosort 2 (*29*) or 2.5 (https://github.com/MouseLand/Kilosort). Clusters identified by Kilosort were manually curated in Phy (https://github.com/cortex-lab/phy). To assess if activity reflected single-unit activity, inter-spike interval, waveform shape, firing rates, activity change across channels, were all examined. Neurons were also assessed for potential merges with similar nearby clusters, and for potential splitting out of noise/other neurons. Neurons were only kept for analysis if the pattern of activity was confidently identified as neuron activity, and not noise or multi-unit activity. Accepted neurons were finally screened using Matlab and only kept if consistently recorded throughout all trials in the session recorded in, any neurons that showed clear drop offs or loss of recordings were discarded. 52% of neurons accepted in Phy passed Matlab screening.

Analysis

Matlab was used to extract, collate, and analyze the single-unit data and behaviour timestamp events. Fear was measured by suppression of rewarded nose poking (baseline poke rate – cue poke rate)/(baseline poke rate + cue poke rate). Perceptually uniform color maps were used to prevent visual distortion of the data (*30*). K-means clustering was performed by systematically varying the number of clusters and examining the output for over/under clustering. Single-unit and population firing analyses utilized k-means clustering, principal components analysis, linear regression combined with iterative shuffling. Complete descriptions of firing analyses provided in supplement.