

NEUROSCIENCE

Facial expressions of emotion states and their neuronal correlates in mice

Nejc Dolensek^{1,2}, Daniel A. Gehrlach^{1,3}, Alexandra S. Klein^{1,3}, Nadine Gogolla^{1*}

Understanding the neurobiological underpinnings of emotion relies on objective readouts of the emotional state of an individual, which remains a major challenge especially in animal models. We found that mice exhibit stereotyped facial expressions in response to emotionally salient events, as well as upon targeted manipulations in emotion-relevant neuronal circuits. Facial expressions were classified into distinct categories using machine learning and reflected the changing intrinsic value of the same sensory stimulus encountered under different homeostatic or affective conditions. Facial expressions revealed emotion features such as intensity, valence, and persistence. Two-photon imaging uncovered insular cortical neuron activity that correlated with specific facial expressions and may encode distinct emotions. Facial expressions thus provide a means to infer emotion states and their neuronal correlates in mice.

Emotions are patterns of behavioral, hormonal, and autonomic responses aimed at promoting survival. Emotions result from brain states that reflect the dynamic integration of external cues, bodily signals, and cognitive processes (1–5). Although emotions have been subject to intensive research efforts in neuroscience, psychology, and philosophy (1, 4, 6, 7), we still lack a mechanistic understanding of how emotions arise in neuronal circuits (3, 4, 8, 9). The functional dissection and causal interrogation of the neuronal circuit underpinnings of emotion rely on research in animal models. However, whether animals experience emotions similar to those of humans and how to best define or investigate emotions are still matters of controversy (3, 5, 8–10). Although most researchers would agree that externally observable behaviors indicate that forms of evolutionarily conserved “emotion states” exist across species (1, 3, 5), investigating emotions using modern neuroscientific tools has been hindered by a lack of rapid and precise readouts of emotion states in model organisms, such as mice (3).

In humans and monkeys, facial expressions have been proposed to provide universal indicators of emotions (11, 12). Rodents may also use their orofacial musculature to signal longer-lasting internal states (13–15). We asked whether mice reacted to emotionally salient stimuli with stereotyped facial expressions and whether these reflect core emotion properties, such as intensity, valence, flexibility, and persistence (3, 4). We then investigated neuronal correlates of inferred emotion states in the insular cortex, an area of the

brain that in humans has been implicated in subjective affective experiences (16, 17).

To study facial expressions, we exposed mice to a diverse set of sensory stimuli that can be assumed to trigger changes in emotion state. In addition to these triggers, we also monitored spontaneous behavioral expressions of emotion states, such as the exhibition of established fear behaviors. These “emotion events” of different types therefore included painful tail shocks, sweet sucrose, bitter quinine, and lithium chloride injections, which induce visceral malaise (14, 18), as well as freezing and escape behaviors (see methods). We video monitored the faces of head-fixed mice (Fig. 1A and fig. S1, A and B). Mice reacted to each emotion event with a noticeable facial movement visible to naïve human observers (Fig. 1B, fig. S2A, and movie S1). However, the valence or type of the underlying emotion event was not intuitively recognizable (fig. S2, B and C) and required extensive experience (Fig. 1B).

To achieve objective and temporally precise classification of facial expressions we used machine vision. We chose “histogram of oriented gradients” (HOG) (19) descriptors to represent the statistics of local image features in a standardized way and provide one numerical vector for each video frame (see materials and methods for advantages of the HOG method). This allowed us to compare facial expressions of mice reacting to emotion events quantitatively through comparison of their corresponding HOG descriptors.

We first assessed the facial expressions resulting from each type of emotion event separately by comparing all video frames collected in the vicinity (before and after) of three repetitions of the same event in individual mice. Pairwise correlations of all frames in these clips rendered two discrete clusters of highly similar facial expressions: One cluster belonged to the pre-event epochs, and the second cluster belonged to the epochs during or immediately after the event (Fig. 1, C and D). No

distinct clusters and thus no consistent change in facial expressions were detected when frames were selected in the same temporal sequence but from mice recorded during a baseline period (see “neutral” condition, Fig. 1D, top).

Next, we examined whether facial expressions were specific to the underlying emotion and visualized frames from all of the emotion events using t-distributed stochastic neighbor embedding (t-SNE). We observed a clean separation into discrete frame clusters for each event type within individual mice, suggesting emotion-specific facial expressions (Fig. 1E and fig. S3).

To test whether the underlying emotion event in any given mouse could be predicted solely from its facial expressions, we trained a random forest classifier (see materials and methods). The decoder could predict each underlying emotion event across different mice reaching accuracies >90%. Performance dropped on average below 15% if the decoder was trained on temporally shuffled data (Fig. 1F, fig. S4, and table S1).

These results raised the question of whether the observed expressions may reflect separate basic emotion states, similar to emotion categories in humans (7, 10). We collected the most characteristic video frames following each type of emotion event separately and averaged the corresponding HOG vectors into a single descriptor (Fig. 2A and materials and methods), which we termed “emotion prototype.” We constructed prototypical HOG descriptors assuming the following event \approx emotion state contingencies: quinine \approx *disgust*, sucrose \approx *pleasure*, tail shock \approx *pain*, lithium chloride \approx *malaise*, escape \approx *active fear*, and freezing \approx *passive fear*.

We first tested the sufficiency of the prototypes to capture the characteristics of the distinct facial expressions across individuals (Fig. 2, B and C, fig. S5, and table S1). We measured the similarity of facial expressions to the emotion prototypes and, indeed, each single prototype was specific to only one emotion state, except for the active fear prototype, which resembled facial expressions evoked by bitter, pain, and escape and may thus capture features of diverse emotion states (Fig. 2C). Comparing each frame of any video sequence across time to an emotion prototype captured the dynamics of facial expressions at high resolution (fig. S6 and movie S2).

Although our results so far suggested that facial expressions may relate to internal emotion states, an alternative explanation could be that facial expressions are stereotyped, reflex-like reactions. We therefore aimed to test whether facial expressions reflected fundamental features of emotions (3, 4), such as intensity, valence, generalization, flexibility, and persistence (Fig. 2, D to G).

Scalability refers to the observation that emotions vary by intensity (3, 5). We thus

¹Circuits for Emotion Research Group, Max Planck Institute of Neurobiology, Am Klopferspitz 18, 82152 Martinsried, Germany. ²Graduate School of Systemic Neurosciences, Ludwig-Maximilians University Munich, Germany.

³International Max-Planck Research School for Molecular Life Sciences, Munich, Germany.

*Corresponding author. Email: ngogolla@neuro.mpg.de

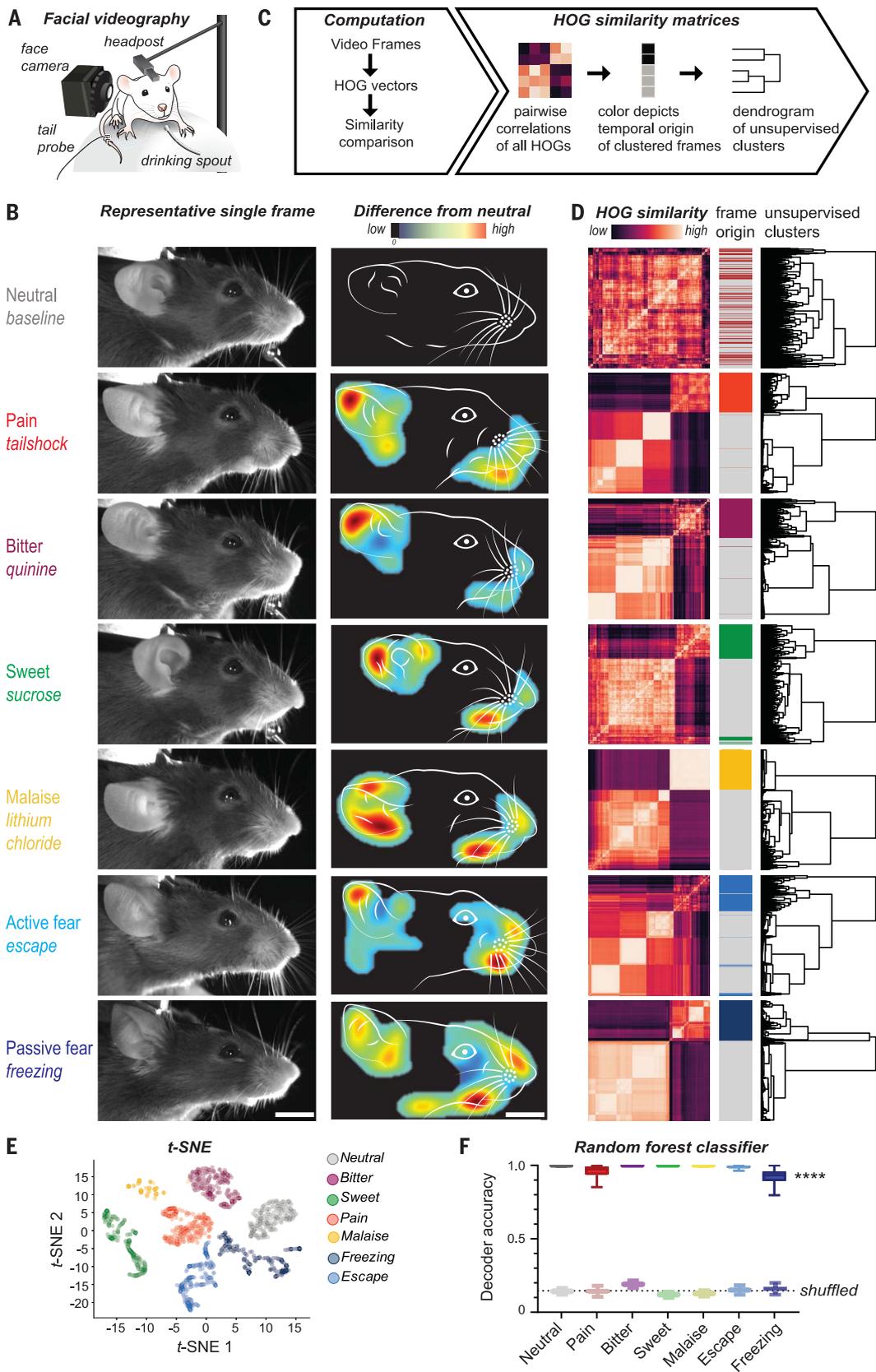


Fig. 1. Emotion-driven facial expressions in mice. (A) Facial videography setup. (B) (Left) Single representative video frames from individual mice captured during baseline (top) or upon different emotion events to illustrate characteristic changes. Images derived from $N = 2$ mice. Similar facial expressions were observed in all animals reported here. (Right) Line drawing of faces from the same frames. Heat-map overlays denote areas of largest difference compared to the neutral expression. Scale bar: 6 mm. (C) Computational strategy to compare facial expressions. (D) Similarity matrices containing pairwise similarity coefficients for all frames obtained in the vicinity of three events for each condition within one animal. To the right, post hoc temporal assignments for each frame are shown in color during the event and in gray before each event. Dendrograms represent hierarchical clustering. (E) t-SNE visualization of frames obtained from all emotion events in an individual mouse. (F) A random forest classifier reliably predicts and distinguishes between all event-related facial expressions. The classifier reaches high decoding accuracies (neutral, bitter, and sweet: $99 \pm 1\%$; pain: $96 \pm 5\%$; freezing: $92 \pm 7\%$; malaise and escape: $99 \pm 2\%$). Decoder performance dropped if the decoder was trained on temporally shuffled data (neutral: $14 \pm 1\%$; bitter: $19 \pm 1\%$; sweet $12 \pm 1\%$; malaise: $13 \pm 1\%$; pain: $14 \pm 2\%$; freezing: $16 \pm 2\%$; escape: $15 \pm 1\%$). Mann-Whitney test revealed a significant ($****P < 0.0001$) difference in classifier's prediction performance between the real and shuffled data for each single facial expression.

varied the stimulus strength and quantified the similarity of the resulting facial expressions to our prototypes. The similarity to prototypical descriptors increased significantly and in a graded manner when the strength of tail

shocks, or the concentration of sucrose or quinine solutions, increased (Fig. 2D and table S1), although the sequence of stimulation did not influence the facial expression intensity at the chosen intertrial intervals (fig. S7, A and B).

Another property of emotions is their valence—namely, they are experienced as good or bad in humans and trigger approach or retreat in animals (3, 5, 14, 18). Salt is appetitive for rodents at low concentrations but aversive at high

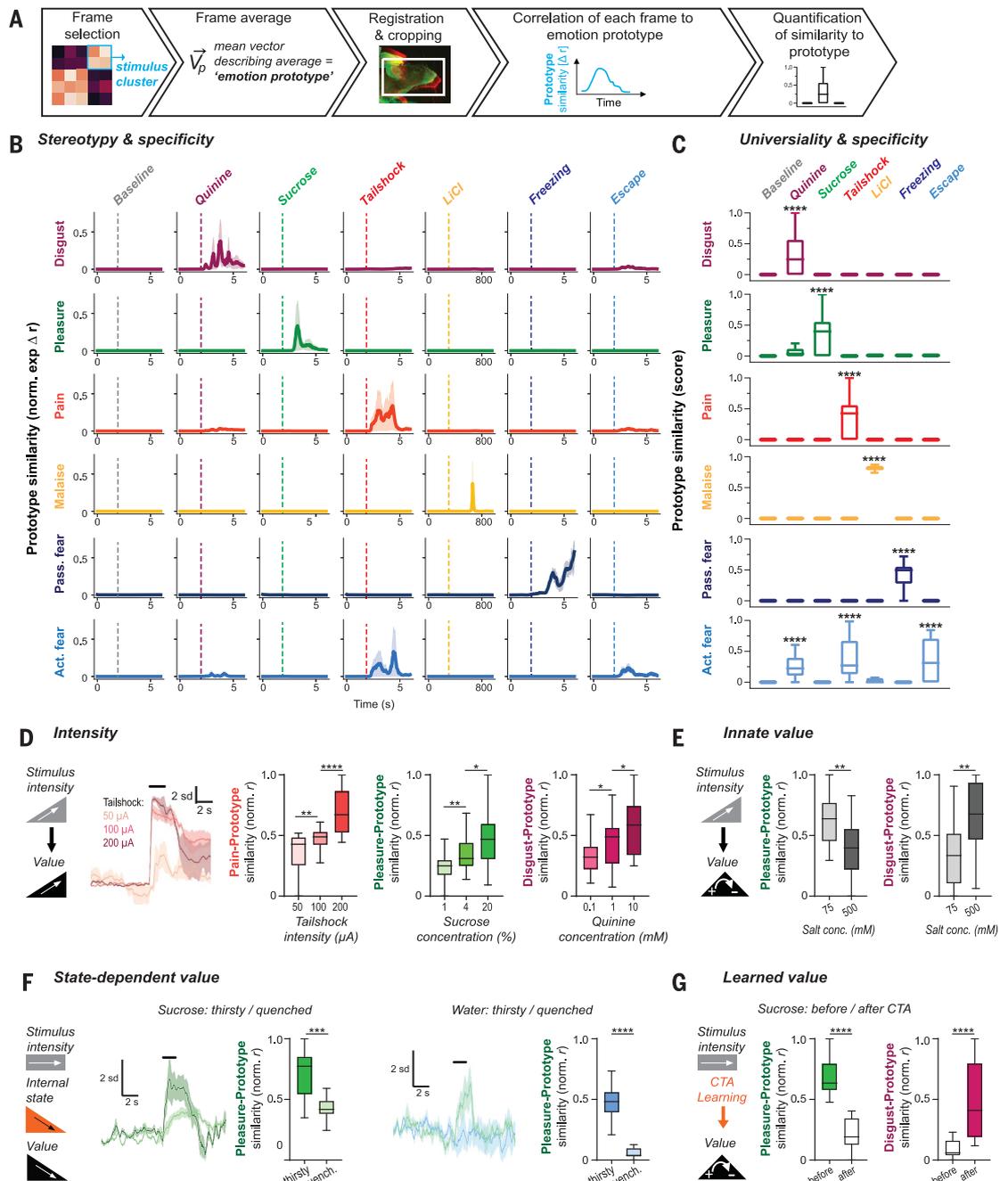
Fig. 2. Facial expressions reflect core features of emotion states.

(A) Schematic of emotion prototype creation. (B) Similarities of facial expressions for each event type (three occurrences each) in one exemplary mouse to each emotion prototype.

(C) Prototypes are valid and specific universally across mice. To calculate a similarity score, data from $N = 9$ mice and $n = 27$ trials per stimulus were averaged, then min-max normalized; the highest similarity value was set to 1, and the maximal baseline value and negative values were set to 0. Facial expressions were highly experience specific [ordinary one-way analysis of variance, **** $P < 0.0001$, Dunnett's post hoc comparisons revealed significant differences (**** $P < 0.0001$) to the neutral condition only for the event matching the prototype, except for escape which carried components of pain and disgust.]

(D) Sensory stimuli of increasing strength elicit more intense facial expressions. (Left) Example traces of face similarities to the pain prototype in one example mouse experiencing increasingly strong tail shocks. To the right, box-and-whisker plots quantifying the facial expression similarity to the pain prototype upon increasing tail shock intensities ($N = 9$ mice, $n = 27$ trials per intensity); the pleasure similarity upon drinking solutions of increasing sucrose content ($N = 9$ mice, $n = 27$ trials per concentration); and disgust similarity upon drinking solutions of increasing quinine content ($N = 10$ mice, $n = 30$ trials per concentration).

(E) Drinking solutions of low salt content (75 mM) evoke pleasure-like facial expression (left) but little disgust-like facial expressions (right). The inverse pattern was observed upon drinking solutions with high salt content (500 mM). $N = 5$ mice, $n = 15$ trials per concentration. (F) Facial expressions reveal the changing affect upon experiencing sucrose or water in either thirsty or quenched states. $N = 5$ mice, $n = 15$ trials per state. (G) Facial expressions reveal associative



aversion learning. Mice expressed highly pleasurable and low disgust facial expressions when drinking sucrose solutions before CTA. After CTA, mice exhibited disgusted facial expressions and low pleasure when drinking sucrose. $N = 5$ mice, $n = 15$ trials per timepoint. In all panels: * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P < 0.0001$, two-tailed Mann-Whitney tests. Box-and-whisker plots in the style of Tukey containing trial averages. Line graphs are z-scored face similarities normalized to the 2 s preceding the stimulus, averaged across three trials in a single animal. Shaded areas are SEM.

concentrations. Facial expressions reflected the innate valence of salt at different concentrations, because salt at low concentration elicited facial expressions of high similarity to our prototypical “pleasure” facial expression and weak similarity to our “disgust” prototype, whereas the opposite was observed for high salt concentrations (Fig. 2E and table S1). Facial expressions are thus decoupled from the underlying stimulus and generalize between different sensory experiences. Both sucrose and low-concentration salt solution elicited pleasure-like expres-

sions, whereas quinine and high-concentration salt solution both evoked disgust.

Emotions reflect an integrated account of external and internal information (3, 9) and are thus flexible. We next varied the internal state of the animal but kept the stimulus constant. When mice drank an identically concentrated sucrose solution or water in either thirsty or quenched states, both liquids elicited significantly stronger pleasure-like facial expressions when mice were thirsty than when they were quenched (Fig. 2F and table S1).

Emotions are thought to arise from predictions about how internal or external events may affect the well-being of the individual (or the well-being of closely related conspecifics) (1, 9, 10). These predictions can depend on the innate or learnt value of stimuli. We already saw how the innate value of salt depended on its concentration. Would learning affect facial expressions in a similar way? We exposed mice to sucrose solution and then injected them with malaise-inducing lithium chloride to induce conditioned taste aversion (CTA). Sucrose

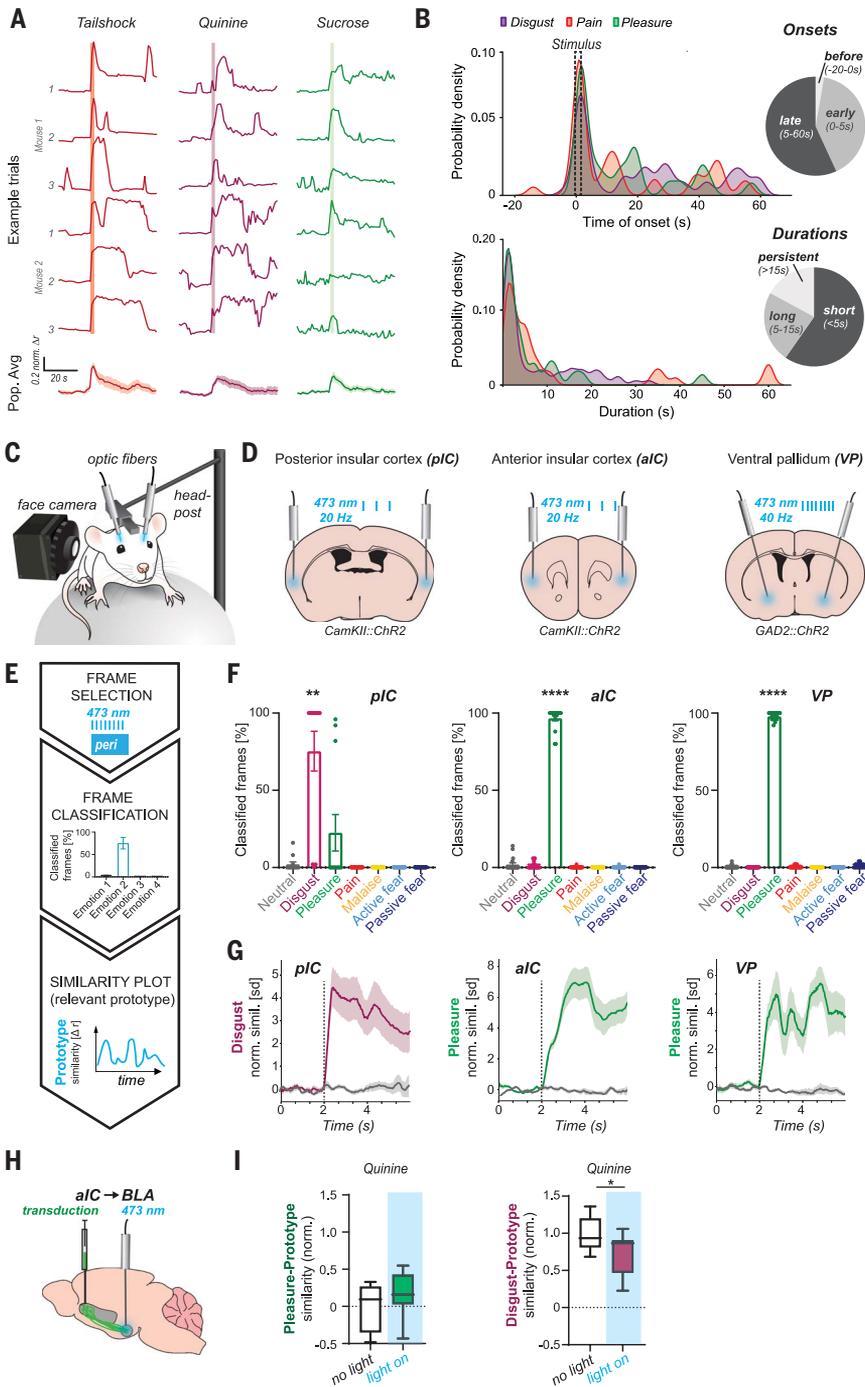
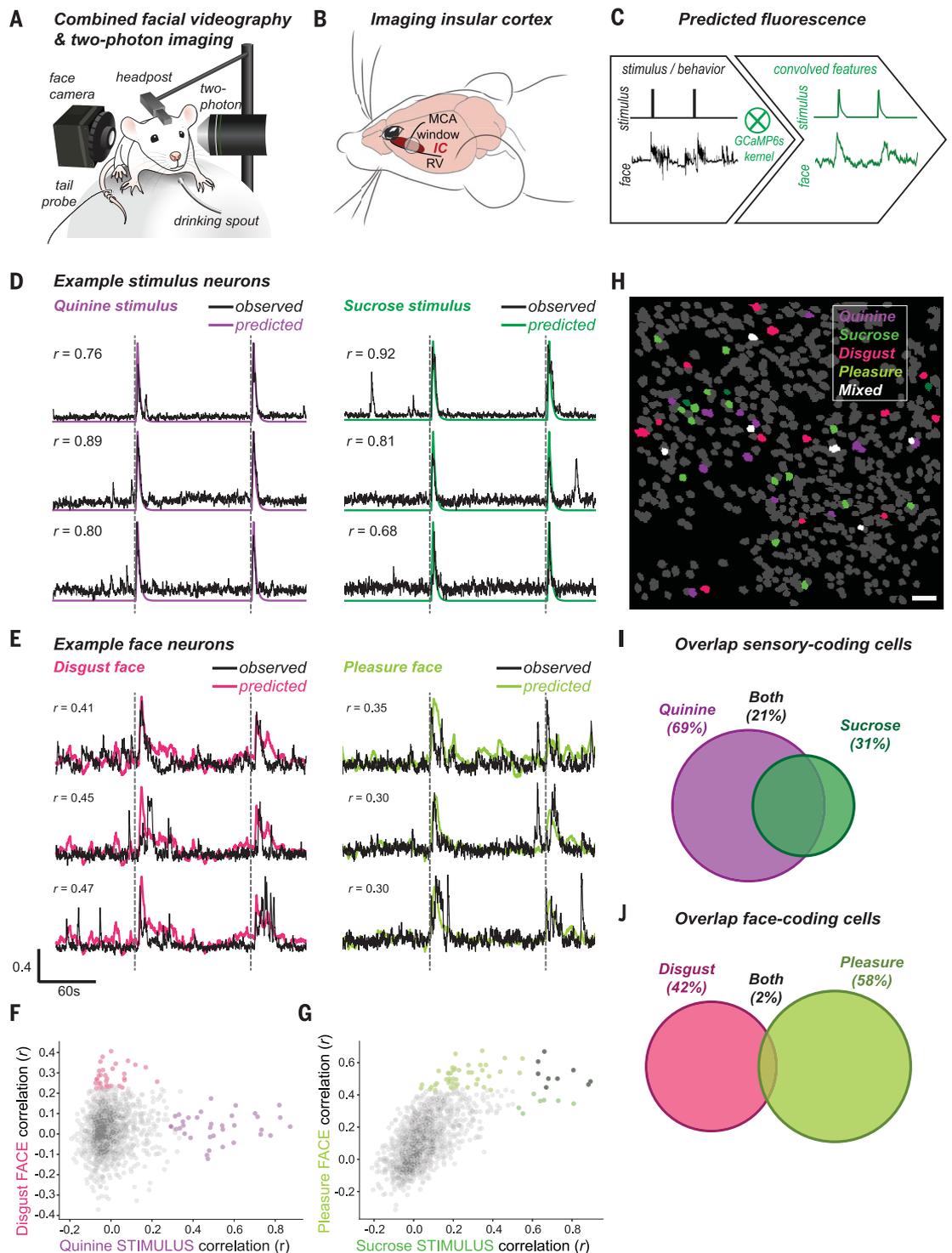


Fig. 3. Facial expressions are variable and associated with internal brain states.

(A) Similarity traces (1-s binned) for each relevant emotion prototype (tail shock, pain prototype; quinine, disgust prototype; sucrose, pleasure prototype). (Top) Individual event-triggered facial expression traces exhibit great variability within the same individual and across mice. (Bottom) Population average (pain and sucrose $N = 9$ animals, $n = 27$ trials; quinine $N = 10$ animals, $n = 30$ trials). Shaded area: 95% confidence interval. (B) Quantification of facial expression onsets (top) and durations (bottom). Probability density is based on kernel density estimates. (C) Experimental approach for combined facial videography and optogenetic circuit manipulations to elicit changes in internal brain states. (D) Optogenetic stimulation sites in the posterior insular cortex (pIC), anterior insular cortex (aIC), and ventral pallidum (VP). (E) Experimental strategy to determine the nature of the optogenetically evoked facial expressions and their description. (F) Individual frames for each optogenetic stimulation epoch were individually classified. For each emotion, the average fraction of classified frames was then plotted per trial (pIC, $n = 12$ trials, $N = 4$ mice; aIC and VP, $n = 18$ trials, $N = 6$ mice). One sample Wilcoxon test revealed significantly higher detection values than random (14.3%) only for one emotion for each optogenetic condition: disgust for pIC and pleasure for aIC and VP (**** $P < 0.0001$). (G) Plot of the normalized similarity (Pearson’s r) for all pre- and peri-event frames against the prototype as suggested by the classifier (dashed line indicates stimulus onset). Lines are mean z-scored face similarities across all trials (as above) with shaded areas representing the SEM. Colored lines from animals expressing ChR2 (channelrhodopsin-2), gray lines from control animals expressing eYFP (enhanced yellow fluorescent protein). (H) Optogenetic strategy to activate the aIC→BLA pathway. (I) Animals were exposed to quinine for 2 s under control (“no light”) and optogenetic activation (“light on”) of the aIC→BLA pathway. $n = 9$ trials from $N = 3$ mice. Similarities were normalized so that during no-light conditions, the mean value for pleasure = 0 and mean value for disgust = 1 in order to reveal the changes from the previously established baseline values.

Fig. 4. Neuronal correlates of emotion state in the posterior insular cortex. (A) Illustration of combined facial videography with awake two-photon calcium imaging. (B) Schematic of the chronic window implant above the posterior insular cortex (IC, red) with respect to major blood vessels: medial cerebral artery (MCA) and rhinal vein (RV). (C) Schematic of neuronal activity prediction through stimulus and face convolution with GCaMP6s kernel. (D and E) Representative normalized fluorescence traces (black) overlaid with predicted stimulus or facial expression traces (colored). R values are Pearson's r for the correlation between normalized fluorescence and the overlaid convolved trace. (F) Scatter plot containing 1198 neurons from two animals experiencing quinine, plotted on the basis of their correlation to the convolved stimulus trace (quinine) and convolved face similarity trace (disgust prototype) for three stimulus presentations. A subset of neurons correlated strongly to the disgust similarity trace is labeled pink. A subset of neurons correlated strongly to the quinine stimulus trace is colored purple (for thresholds, see materials and methods). (G) Same as (F), but with sucrose stimulus. Neurons most strongly correlated to the pleasurable facial expression are labeled light green, neurons most strongly correlated to sucrose stimulus are in dark green, and the subset of neurons highly correlated to both are colored black (for thresholds, see materials and methods). (H) An example field of view from one animal with labeled regions of interest (ROIs) (gray circular shapes). Neurons, as identified and labeled in (F) and (G), are overlaid with the appropriate color. White ROIs indicate neurons with mixed coding properties (mostly multisensory neurons). (I and J) Venn diagrams representing the overlap in coding properties between sensory-coding cells (I) and face-coding cells (J). Scale bar: 100 μm .



before CTA learning elicited pleasure but not disgust. After CTA learning, mice displayed disgusted facial expressions in response to sucrose and thus their expressions reflected the learnt change in subjective value of sucrose (Fig. 2G and table S1).

Emotions are thought to reflect complex internal brain states. Because we cannot control all emotion-relevant information streams, one would hypothesize that even under identical stimulus conditions, the triggered emotion state should vary. We therefore analyzed

the variability of stimulus-triggered facial expressions. Within the same mouse but also across different mice, repeating the same stimulus elicited facial expressions that varied in intensity, onset, and duration (Fig. 3, A and B). Facial expressions could wane and

spontaneously reappear, possibly reflecting dynamic fluctuations in the underlying emotion state (Fig. 3A). Although the great majority of stimulus presentations resulted in immediate facial expressions (~90% of stimuli evoked facial expressions within 5s of stimulus onset), a considerable number of overall facial expressions occurred late after the stimulus (>5 s after stimulus start). Similarly, the duration of facial expressions was highly variable. Most facial expressions triggered by 2-s-long sensory stimuli lasted for less than 5 s (~60%); however, a substantial fraction of facial expressions lasted for relatively long periods (5- to 15-s duration, ~23%), or even persisted for more than 15 s (~17%) (Fig. 3B).

Direct brain stimulations can evoke specific emotions (20, 21). We used optogenetics to test whether manipulating activity in emotion-relevant neuronal circuits could drive facial expressions (Fig. 3, C and D). We activated subregions and specific projections of the insular cortex (IC) that have been shown in humans and animals to evoke emotional sensations and behaviors (20, 22–25). Furthermore, we manipulated the γ -aminobutyric acid-releasing neurons in the ventral pallidum (VP) that process rewarding properties of pleasant stimuli (26) (Fig. 3D). Each region-specific optogenetic manipulation evoked strong facial expressions (fig. S8 and movie S3). To analyze whether the evoked facial expressions would fall into our previously created emotion-state categories, we used the same random forest classifier as in Fig. 1F and categorized all frames during the optogenetic stimulations (Fig. 3, E and F). For each of these three manipulations, the classifier identified one specific emotion to be displayed—namely, pleasure—for the anterior IC and VP, but disgust for the posterior IC stimulations (Fig. 3F and table S1). When we compared the optogenetically evoked facial expressions to our emotion prototypes, we found a similar temporal build-up and persistence of the facial expressions to those triggered externally (Fig. 3G and movie S3). Projections from the insular cortex to the amygdala can influence the emotional value of tastants (25). Indeed, in agreement with this earlier report, the activation of the anterior IC→basolateral amygdala (aIC→BLA) pathway during the exposure to quinine attenuated the expression of disgust (Fig. 3, H and I).

Our data so far suggest that facial expressions are sensitive reflections of internal emotion states, which correspond to brain states. Therefore, we assumed that facial expressions should have neuronal correlates in emotion-relevant brain regions. The insular cortex is a critical brain region for emotional experience and behavior (16, 17, 20–24). We combined facial videography with two-photon calcium imaging in the posterior IC (pIC) to search for neuronal correlates of facial expressions (Fig. 4, A and B,

and fig. S9). We identified single neurons that reliably encoded sensory stimuli in the pIC (Fig. 4, C to G), consistent with previous studies (22, 27). We also identified neurons that exhibited strong correlations to the facial expression dynamics and only low correlations with the stimuli (Fig. 4, D to G). Indeed, these “face” neurons captured the characteristic persistence and spontaneity of the facial expression. Although a substantial fraction of stimulus neurons was multisensory, face-responsive neurons were highly segregated and exhibited almost no overlap.

In this study, we have identified facial expressions as reliable indicators of emotion states and their neuronal correlates in mice. But why do mice exhibit facial expressions? Charles Darwin suggested that facial expressions reveal affective processes across species, implying an evolutionarily conserved function of these behaviors (1). Though often discussed in the context of social communication, facial expressions may have evolved first as parts of emotional action programs, preparing for motor behaviors and adapting sensory acquisition to changes in the internal or external milieu (2, 28, 29). Indeed, head-fixed mice, which do not socially interact, consistently respond to emotionally salient events with stereotyped facial expressions. Although the value of facial expressions for uncovering emotional processes in humans remains controversial (30), this may be partially due to the volitional control that humans exert over emotions and their expression. It would therefore be interesting to examine how facial expressions are modified by the presence of conspecifics in mice.

Direct observation of facial expressions is possible in quasi-real time (fig. S10) and allows for the mechanistic investigation of the neural underpinnings of emotions in mice. Correlation of emotional facial expressions with neuronal activity recordings and closed-loop manipulations are promising approaches to search for and test the causal role of the neuronal substrates of basic emotional building blocks, such as intensity, valence, and persistence.

Our data suggest that facial expressions can be classified into different basic categories. An important question for future studies may be to what degree emotion states are dimensional or categorical states at the level of not only behavioral expressions but also the underlying brain circuitries. The relatively simple implementation of HOG feature descriptors may become a useful addition to studying emotional facial or postural expressions in other laboratory animals, such as rats, shrews, lemurs, and monkeys. It may also help in identifying unknown, species-specific emotion states and assist in moving toward a more universal and evolutionarily based definition and understanding of emotions and their neural underpinnings across species.

REFERENCES AND NOTES

1. C. Darwin, *The Expression of the Emotions in Man and Animals* (London, Murray, 1872).
2. A. Damasio, G. B. Carvalho, *Nat. Rev. Neurosci.* **14**, 143–152 (2013).
3. D. J. Anderson, R. Adolphs, *Cell* **157**, 187–200 (2014).
4. Ralph Adolphs & David J. Anderson, *The Neuroscience of Emotion: A New Synthesis* (Princeton Univ. Press, 2018).
5. R. Adolphs, *Soc. Cogn. Affect. Neurosci.* **12**, 24–31 (2017).
6. J. A. Russell, *J. Pers. Soc. Psychol.* **39**, 1161–1178 (1980).
7. P. Ekman, *Cogn. Emotion* **6**, 169–200 (1992).
8. J. LeDoux, *Neuron* **73**, 653–676 (2012).
9. L. F. Barrett, A. B. Satpute, *Neurosci. Lett.* **693**, 9–18 (2019).
10. J. Panksepp, *Neurosci. Biobehav. Rev.* **35**, 1791–1804 (2011).
11. P. Ekman, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **335**, 63–69 (1992).
12. L. A. Parr, B. M. Waller, J. Fugate, *Curr. Opin. Neurobiol.* **15**, 716–720 (2005).
13. D. J. Langford et al., *Nat. Methods* **7**, 447–449 (2010).
14. A. Faure, J. M. Richard, K. C. Berridge, *PLOS ONE* **5**, e11223 (2010).
15. K. Finlayson, J. F. Lampe, S. Hintze, H. Würbel, L. Melotti, *PLOS ONE* **11**, e0166446 (2016).
16. J. Chikazoe, D. H. Lee, N. Kriegeskorte, A. K. Anderson, *Nat. Neurosci.* **17**, 1114–1122 (2014).
17. N. Gogolla, *Curr. Biol.* **27**, R580–R586 (2017).
18. P. Tovote, J. P. Fadok, A. Lüthi, *Nat. Rev. Neurosci.* **16**, 317–331 (2015).
19. N. Dalal, B. Triggs, in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005* (2005), vol. 1, pp. 886–893.
20. F. Caruana, A. Jezzini, B. Sbriscia-Fioretti, G. Rizzolatti, V. Gallese, *Curr. Biol.* **21**, 195–199 (2011).
21. C. Fausto, *Emot. Rev.* **11**, 27–38 (2019).
22. D. A. Gehlrich et al., *Nat. Neurosci.* **22**, 1424–1437 (2019).
23. J. Yih, D. E. Beam, K. C. R. Fox, J. Parvizi, *Soc. Cogn. Affect. Neurosci.* **14**, 339–351 (2019).
24. Y. Peng et al., *Nature* **527**, 512–515 (2015).
25. L. Wang et al., *Nature* **558**, 127–131 (2018).
26. L. Faget et al., *Nat. Commun.* **9**, 849 (2018).
27. X. Chen, M. Gabitto, Y. Peng, N. J. P. Ryba, C. S. Zuker, *Science* **333**, 1262–1266 (2011).
28. C. Montag, J. Panksepp, *Motiv. Emot.* **40**, 760–766 (2016).
29. J. M. Susskind et al., *Nat. Neurosci.* **11**, 843–850 (2008).
30. L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, S. D. Pollak, *Psychol. Sci. Public Interest* **20**, 1–68 (2019).
31. N. Dolensek, N. Gogolla, *GogollaLab/MouseFacialExpressionAnalysis*, Version 1.0, Zenodo (2020).

ACKNOWLEDGMENTS

We thank members of the Gogolla laboratory, K. Branson, P. Dayan, W. Denk, M. Hübener, E. Mace, D. Mearns, R. Portugues, and A. Sirota for discussions; J. Kuhl (somedonkey.com) for illustrations; and T. Black, F. Lyonnaz, A. Podgornik, and C. Weiland for technical assistance. **Funding:** Supported by the Max-Planck Society, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-2017-STG, grant agreement no. 758448 to N.G.), the Deutsche Forschungsgemeinschaft (SP11665), the German Israeli Foundation (Grant I-1301-418.13/2015), and the ANR-DFG project “SAFENET” (ANR-17-CE37-0021). **Author contributions:** N.G. and N.D. conceived the project and designed the experiments. N.D. performed all experiments and developed and performed all facial expression analysis. D.A.G. and A.S.K. helped with the optogenetic experiments. N.G. wrote the manuscript with assistance from N.D. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. The facial expression analysis code is available on GitHub (<https://github.com/GogollaLab>) and at Zenodo (31). Viruses were packaged at the University of North Carolina (UNC) Vector Core and made available under a material transfer agreement.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6486/89/suppl/DC1
Materials and Methods
Figs. S1 to S10
Table S1
Movies S1 to S3
References (32–41)

[View/request a protocol for this paper from Bio-protocol.](#)

22 October 2019; accepted 11 February 2020
10.1126/science.aaz9468