FACIAL EXPRESSIONS OF EMOTION STATES AND THEIR CORRELATES IN INSULAR CORTEX

Nate Nejc Dolensek

Dissertation at the
Graduate School of Systemic Neurosciences
Ludwig-Maximilians-Universität München

March 2021

Supervisor
Dr. Nadine Gogolla
Max-Planck-Institute of Neurobiology, Martinsried, Germany

First Reviewer: Dr. Nadine Gogolla
Second Reviewer: Prof. Dr. Mark Hübener

Date of Submission: 26.3.2021
Date of Defense : 15.10.2021

TABLE OF CONTENTS

1. SUMMARY

Emotions have long been a source of intrigue across many basic sciences and an intense focus of clinical research, with psychiatric disorders involving emotional dysfunction being among the most prevalent. Unfortunately, we, to this day, lack mechanistic understanding of how emotions arise in neuronal circuits and even the definition of emotion is still a highly disputed topic. One of the major hindrances to the study of emotion has been our lack of understanding of how such internal states link to observable behavior, and while expressions of emotion have been explored for centuries, most famously by Darwin and Ekman, they are still somewhat controversial in humans, with even more unknowns regarding their presence in non-primates.

This thesis consists of three published manuscripts and represents an attempt at making internal emotion states measurable by linking them to objectively observable behavior and then using this entry point to mechanistically explore the underlying neural circuitry.

First manuscript identifies facial expressions as a promising such observable behavior in mice, providing a comprehensive description of facial expressions of emotion and introduces machine vision approaches to objectively quantify them. Crucially, this opens a door to using modern neuroscientific approaches for neuronal observation and perturbation, further identifying cells in insular cortex as accurately reflecting emotional experience on a moment-to-moment basis. Second manuscript significantly expands on the first manuscript, focusing specifically on insular cortex and demonstrates its important role across a whole spectrum of emotional behaviors, further showing that it contains neural representations that might be at the core of emotion states of short and long timescales, which are also associated with facial expressions. Third manuscript summarizes and briefly discusses machine learning approaches useful for studying emotion states in mice and offers a perspective on future work on untangling emotion by using modern approaches to study the brain and behavior.

Taken whole, this thesis introduces tools and discoveries that make emotions more tangible by expanding available approaches for studying them in mice and combines these with deep neural circuit dissection tools available in mouse neuroscience. Finally, this led to significant new insights in understanding how emotion is implemented in the brain.

2. INTRODUCTION

2.1 EMOTION AND ITS DEFINITION

Emotions in humans and animals were famously already studied by Charles Darwin (Darwin & Darwin, 2009), who argued that they are innate, evolutionarily adaptive and exist across species. However, over the last century and even today (Zych & Gogolla, 2021), definition of emotion remains a topic of intense debate across scientific fields. For the purpose of this thesis, I selected the following biological definition of emotion in line with an emotion framework proposed by Anderson and Adolphs (Anderson & Adolphs, 2014): Emotions can be defined as survival promoting patterns of behavioral, autonomic, and hormonal responses. More specifically, emotions are brain states which drive responses to stimuli in a fitness promoting manner and are therefore evolutionarily conserved, including their neural implementations across species.

2.2 HISTORIC PERSPECTIVE ON HUMAN FACIAL EXPRESSION

Ancient Greeks already understood the link between emotions and their expressions. In their theaters, actors wore large clay masks stylized to represent characters' faces and the emotions they were experiencing (Meineck, 2011). For a more immersive experience the actors trained extensively to further convey emotional content via bodily movements and voice.

In 17th century, an influential French painter Charles Le Brun (Fig. 1), building on contemporary ideas by René Descartes, argued for the importance of facial expressions of emotions, then known as passions, in life and art. Le Brun had this to say (Lyons, 2019): *"Ordinarily whatever causes passion in the soul occasions some action in the body. Since it is thus true that most passions of the soul produce bodily actions, we need to know which actions of the body express the passions and also what an action is. Action is nothing more than the movement of some part and change comes about only by a shift in the muscles. [...] If it is true that there is a part where the soul has the most immediate effect, and if that part is the brain, we can also say that the face is the part of the body where the soul most particularly manifests what it feels."* Le Brun further systematically explored similarities between facial features and expressions in people and animals. Unfortunately, much of Le Brun's innovative work served as a foundation for the revival of physiognomy (Hartley, 2001), a now discredited study of person's character by observing facial features, which was further popularized by Swiss pastor and writer Johann Kaspar Lavater and Charles Bell, a Scottish neurologist famous especially for his discovery of motor control of facial musculature via the seventh cranial nerve.

Figure 1: Sketches of passions (facial expressions of emotion) by Charles Le Brun (Charles Le Brun, the Expressions.jpg - Wikimedia Commons, 2011; public domain)

In 19th century, Duchenne de Boulogne (Fig. 2), now recognized as the father of modern neurology and electrophysiology (Hueston & Cuthbertson, 1978; Parent, 2005), set to prove that the premises of physiognomy are false. He believed that face does not reflect the moral character but rather reveals the momentary emotional state of an individual. For this reason, he believed that the fleeting facial expressions cannot be accurately captured by drawing or description but require an alternative objective method of acquisition which is fast and objective – photography. He believed that different sets of facial muscles contracted in different ways based on emotional state, representing a sort of a map or an innate language that can be decoded into a taxonomy of internal states, writing: *"In the face our creator was not concerned with mechanical necessity. He was able in his wisdom or – please pardon this manner of speaking – in pursuing a divine fantasy … to put any particular muscles into action, one alone or several muscles together, when He wished the characteristic signs of the emotions, even the most fleeting, to be written briefly on man's face. Once this language of facial expression was created, it sufficed for Him to give all human beings the instinctive faculty of always expressing their sentiments by contracting the same muscles. This rendered the language universal and immutable."* To study this "language of emotion" de Boulogne's used of electrical stimulators on different muscles of the face in combination with photography and was able to recreate the spectrum of natural human expressions, going as far as to propose a set of 13 primary emotions based on the number of muscles needed to create the associated expression.

Figure 2: Duchenne de Boulogne performing a facial electrostimulation experiment. (Guillaume Duchenne De Boulogne Performing Facial Electrostimulus experiments.jpg - Wikimedia Commons, 2005; public domain.)

De Boulogne's work strongly influenced Charles Darwin (Fig. 3), who approximately a decade later published The Expression of the Emotions in Man and Animals (Darwin & Darwin, 2009), observing that facial expressions of emotion, and therefore emotions, are innate, universal, biologically adaptive and might exist in humans and animals. Darwin further reduced the number of basic emotions to six: happiness, fear, sadness, anger, surprise, and disgust and argued against strong cultural influence on emotion or its expression, observing that expressions were already present in small children. However, many Darwin's conclusions were strongly disputed, and universality of emotion and expression only became generally accepted in the second half of the 20th century, when Paul Ekman and Wallace V. Friesen (Paul Ekman et al., 1969) published a series of studies investigation emotion and expression across cultures, going as far as to establish contact with pre-literate tribes in Papua New Guinea and observing high cross-cultural agreement between emotional labels and linked expressions, proving their innateness and universality. To enable more objective and quantitative study of facial expression and finally further de Boulogne's goal of creating a taxonomy of facial expression, Ekman and Friesen created Facial Action Coding System or FACS (P. Ekman & Friesen, 1978), a formal approach to registering any possible facial expression via visual inspection of contractions of individual facial muscles and linking these sets of muscular contractions to emotions. While initially manual, FACS has recently been computerized and automated (De La Torre et al., 2011).

2.3 FACIAL EXPRESSION OF EMOTION IN ANIMALS

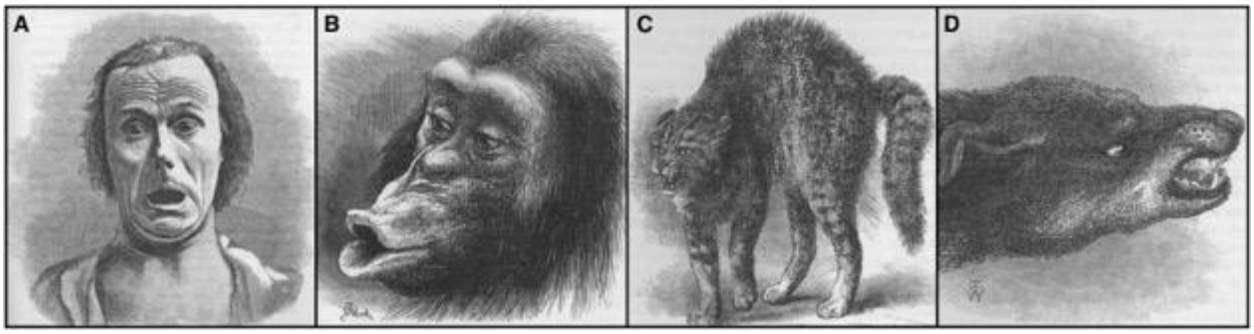


Figure 3: Examples of emotion expressions in a human and animals as observed by Darwin (Darwin & Darwin, 2009)

Facial expressions have received considerably less attention in animal models than in humans, with animal emotions themselves still being relatively controversial (Panksepp, 2011). Facial expressions have been explored in various primates (L. A. Parr et al., 2010; Lisa A. Parr et al., 2005), who have been shown to employ them both to express emotion and as a non-verbal communication tool. In 1970's Grill and Norgren (Grill & Norgren, 1978a, 1978b) demonstrated that rodents express what they called "liking" and "disgust" using orofacial movements in response to delivery of appetitive and aversive tastants (Fig. 4), interestingly reflecting both intrinsic and learned values. Further work by Berridge and colleagues (Berridge, 2018; Berridge & Kringelbach, 2013; Winkielman et al., 2005) identified parallels across expressions of liking and disliking across rodents, primates, and humans, indicating that basic emotions might be shared across mammals and have a similar neural implementation.

Figure 4:Orofacial indicators of liking and disgust exist across mammals (Berridge & Kringelbach, 2013)

More recently, scientists described orofacial indicators of longer lasting pain (Langford et al., 2010; Sotocinal et al., 2011) (Fig. 5), positive affect (Finlayson et al., 2016) and aggression (Defensor et al., 2012) in rodents. Additionally, FACS has been successfully applied to seven animal species, among them horses, dogs and cats (Caeiro et al., 2017; Micheletta et al., 2015; L. A. Parr et al., 2010; Lisa A. Parr et al., 2007; Waller et al., 2013; Wathan et al., 2015), indicating that approaches to analyzing human orofacial behavior might be transferable to other mammals.

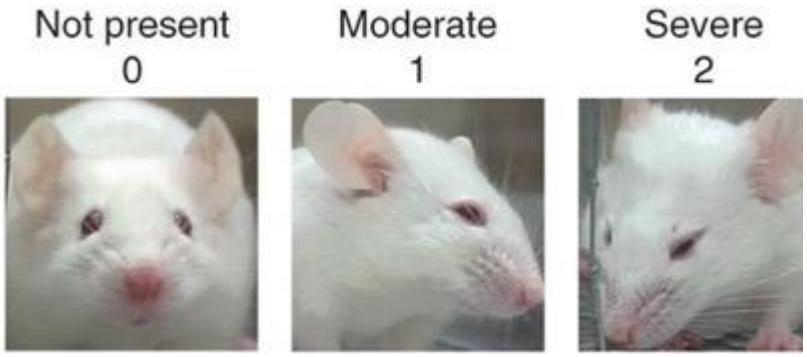


Figure 5: Mouse pain grimace scale (Langford et al., 2010)

Advances in computer capabilities and dissemination of computational analysis, computer vision and machine learning have had a profound effect on neuroscience. Facial expression analysis is not an exception, with systems like FACS now existing in automated, computerized versions (De La Torre et al., 2011). Advances in technology did not solely result in automated versions of previously manual expression scoring tools, but also lead to a large number of alternative human facial expression recognition methods (Samadiani et al., 2019), some relying on modern statistical and machine learning tools to distinguish expressions in an unsupervised, rather than a supervised manner.

2.4 THE INSULAR CORTEX

The insular cortex (IC) is a subdivision of cerebral cortex, in rodents located on the lateral cortical surface, above the rhinal fissure (Gogolla, 2017). In humans, who possess a folded brain, insular cortex is a deeply invaginated cortical area within the temporal sulcus. IC is commonly divided along the rostro-caudal axis into anterior and posterior insula. While this is an anatomical distinction, these sections, although strongly interconnected, differ significantly in their brain-wide connectivity and function (Gehrlach et al., 2020). Furthermore, IC can be divided based on cyto-architecture, into granular (6-layer cortex), dysgranular and agranular (5-layer cortex which lacks layer IV) insula, which roughly follow each other in a dorso-ventral axis.

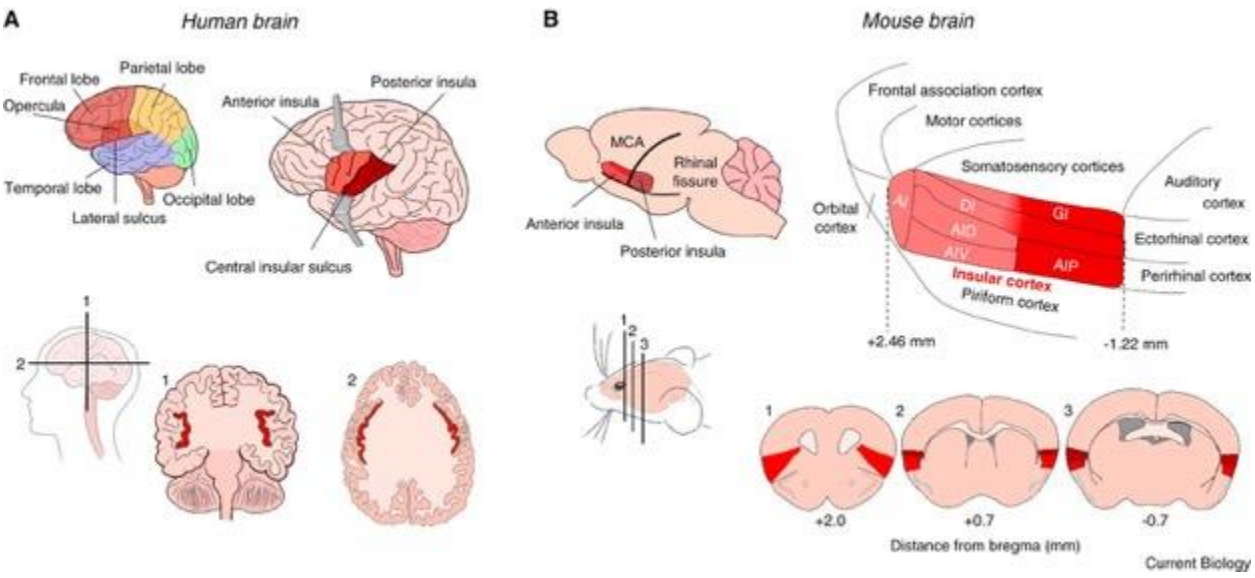


Figure 6: Insular cortex in the human and rodent brain (Gogolla, 2017)

Connectivity wise, IC is a major hub (Gehrlach et al., 2019, 2020; Gogolla, 2017), with strong reciprocal connections to many cortical and subcortical areas, connecting several distinct brain networks, fitting for its role as an important brain center of information integration. IC receives wealth of extra- and intra-sensory information from across modalities, both from sensory cortices and directly from the thalamus. Moreover, parts of insula perform primary sensory cortex functions, for example in gustation and interoception, the sensation of own body,

with sections of insula also being known as primary gustatory and primary interoceptive or visceral cortex. Furthermore, IC has been shown to receive proprioceptive input from facial muscles (Ikenoue et al., 2018; Sato et al., 2017; Tsutsumi et al., 2018) and to project to brain areas known to be involved in emotion and expression of emotion like the parvocellular nucleus of reticular formation, amygdalar nuclei and frontal and motor cortical areas (Gehrlach et al., 2020).

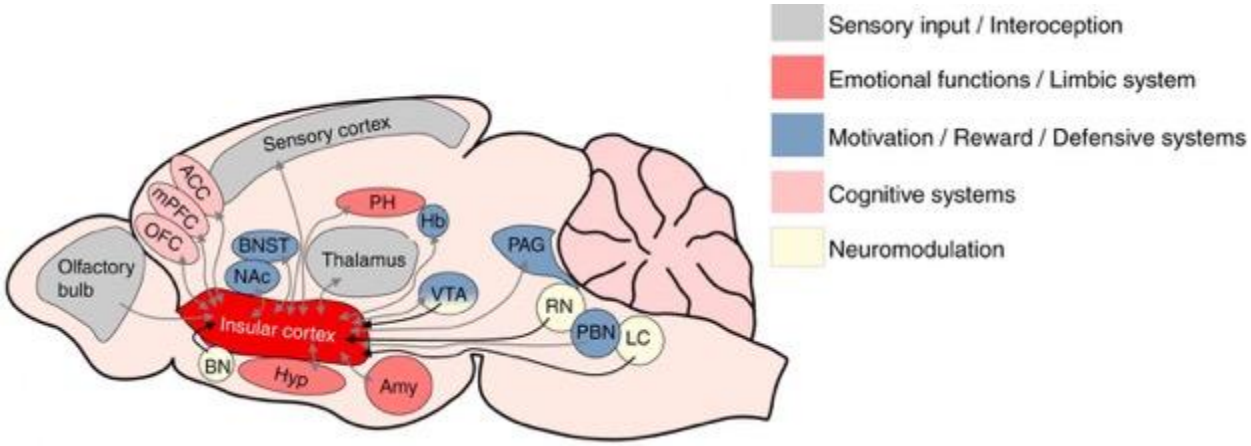


Figure 7: Connectivity of insular cortex (Gogolla, 2017)

Human fMRI studies have long implicated IC in emotion processing and regulation, with changes in IC activity being associated with many disorders of emotion (Gogolla, 2017; Paulus & Stein, 2010; Sliz & Hayley, 2012), for example anxiety disorders. Association between insula and anxiety has been also shown in rodents (Méndez-Ruette et al., 2019). Insula has also been proposed as a detector of salient stimuli of any modality, gated based on the homeostatic state (Livneh et al., 2017). Very interestingly, IC has been shown to activate when observing facial expressions (Lin et al., 2016), lesions of insula in humans have been shown to lead to impairments in facial expression recognition (Phillips et al., 1997; Terasawa et al., 2015) and lesions in rodents have shown a significant decrease in production of grimaces associated with pain (Langford et al., 2010).

2.5 SYNTHESIS – MAKING INVISIBLE, VISIBLE

Modern neuroscience is principally based on the idea that studying the biological substrate for cognition – the nervous system – is sufficient for gaining a complete understanding of its function to the level of its most minute detail. This principle establishes that even what some might consider as entirely subjective experiences, for example, of emotional states, exist as specific patterns of measurable changes in activity of the nervous system. However, as the brain is typically constantly active, it is nontrivial to determine which of the activity corresponds to one or another function, even more so when the function is not easily observable externally, as in internal emotion states. Our ability to accurately determine neuronal function is further limited by the fact that even cutting-edge neurophysiological tools enable observation and manipulation of only small sections of mammalian brain at a time.

Decades ago, Kupferman and Weiss (1978) famously established a framework for rigorously determining mechanistic neural function by formalizing the concept of a "command neuron", a neuron which has a critical function in a normally occurring behavior. They postulated that if a neuron is active during a certain behavior (activity), its artificial activation produces such behavior in absence of any external stimuli (sufficiency) and its removal prevents the behavior from forming even in response to relevant stimuli (necessity) such neuron can be considered a command neuron. At first glance, one might consider such a framework unsuitable for studying internal cognitive phenomena like emotion as they are not inherently linked to observable behavior. However, if accurate behavioral correlates of emotion can be identified in a model organism where neural activity can be precisely recorded and perturbed at scale, then a door to identifying command neurons of emotion might be open. Existing work already made significant strides toward understanding the neural substrate of emotion using such an approach by recording and manipulating neural activity in mice and observing relatively unspecific behaviors like avoidance, approach and freezing, mainly in freely moving mice (Zych & Gogolla, 2021; Anderson & Adolphs, 2014) Importantly, the goal of studying emotion is not simply to understand emotional behavior but to capture and decode the actual ethereal subjective

emotional experience on a moment to moment basis. Identifying a behavior specific to emotion and reflective of its properties and qualities (Anderson & Adolphs, 2014) would open a door to significant strides to be made in our understanding of emotional function.

3. FIRST MANUSCRIPT

From:

Dolensek, N., Gehrlach, D. A., Klein, A. S., & Gogolla, N. (2020). Facial expressions of emotion states and their neuronal correlates in mice. *Science, 368*(6486). https://doi.org/10.1126/science.aaz9468

Reprinted with permission from AAAS.

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 ≈ *emotion state* contingencies: quinine ≈ *disgust*, sucrose ≈ *pleasure*, tail shock ≈ *pain*, lithium chloride ≈ *malaise*, escape ≈ *active fear*, and freezing ≈ *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

[1]Circuits for Emotion Research Group, Max Planck Institute of Neurobiology, Am Klopferspitz 18, 82152 Martinsried, Germany. [2]Graduate School of Systemic Neurosciences, Ludwig-Maximilians University Munich, Germany. [3]International Max-Planck Research School for Molecular Life Sciences, Munich, Germany.
*Corresponding author. Email: ngogolla@neuro.mpg.de

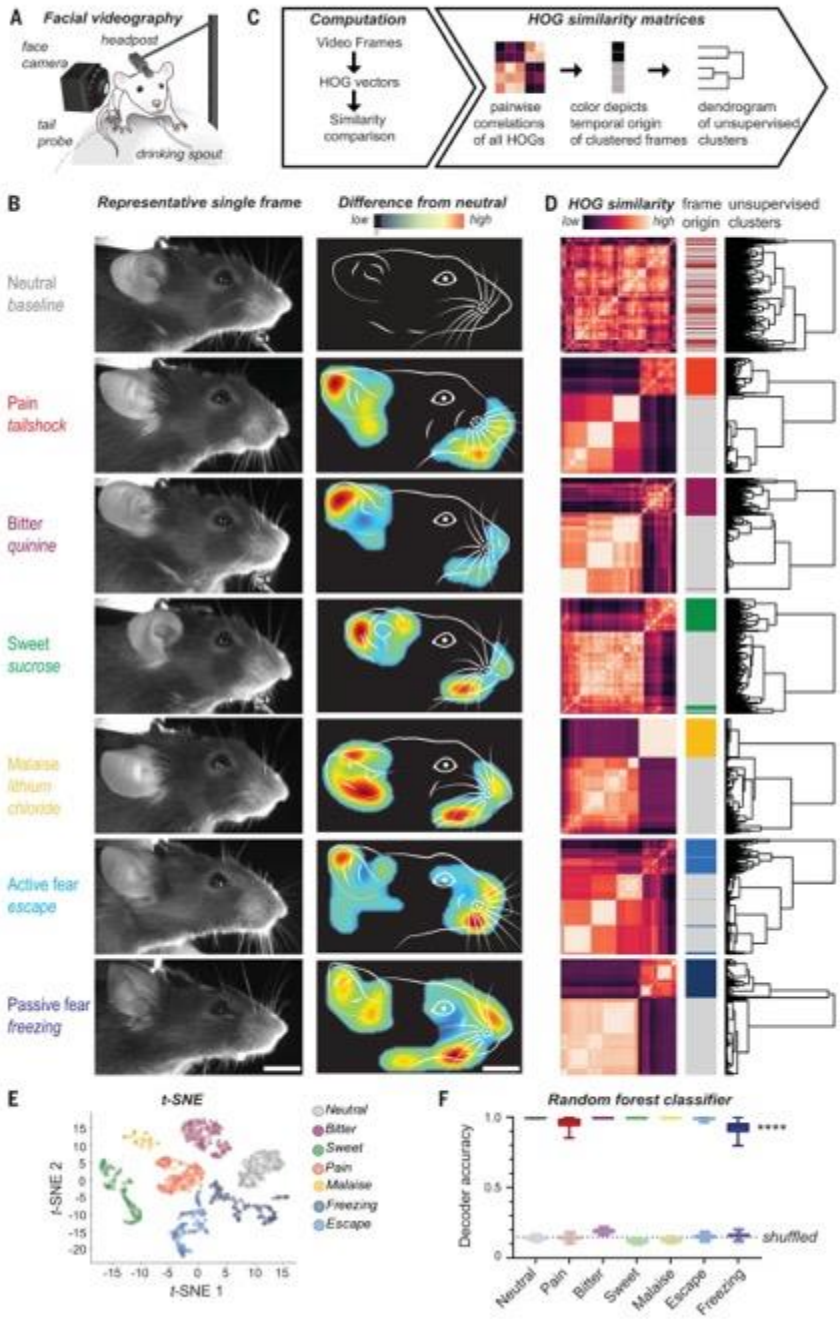


Fig. 1. Emotion-driven facial expressions in mice. (**A**) Facial videography setup. (**B**) (Left) Single representative video frames from individual mice captured during baseline (top) or upon different emotion events to illustrate characteristic changes. Images derived from $N = 2$ mice. Similar facial 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 ± 1%; pain: 96 ± 5%; freezing: 92 ± 7%; malaise and escape: 99 ± 2%). Decoder performance dropped if the decoder was trained on temporally shuffled data (neutral: 14 ± 1%; bitter: 19 ± 1%; sweet 12 ± 1%; malaise: 13 ± 1%; pain: 14 ± 2%; freezing: 16 ± 2%; escape: 15 ± 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

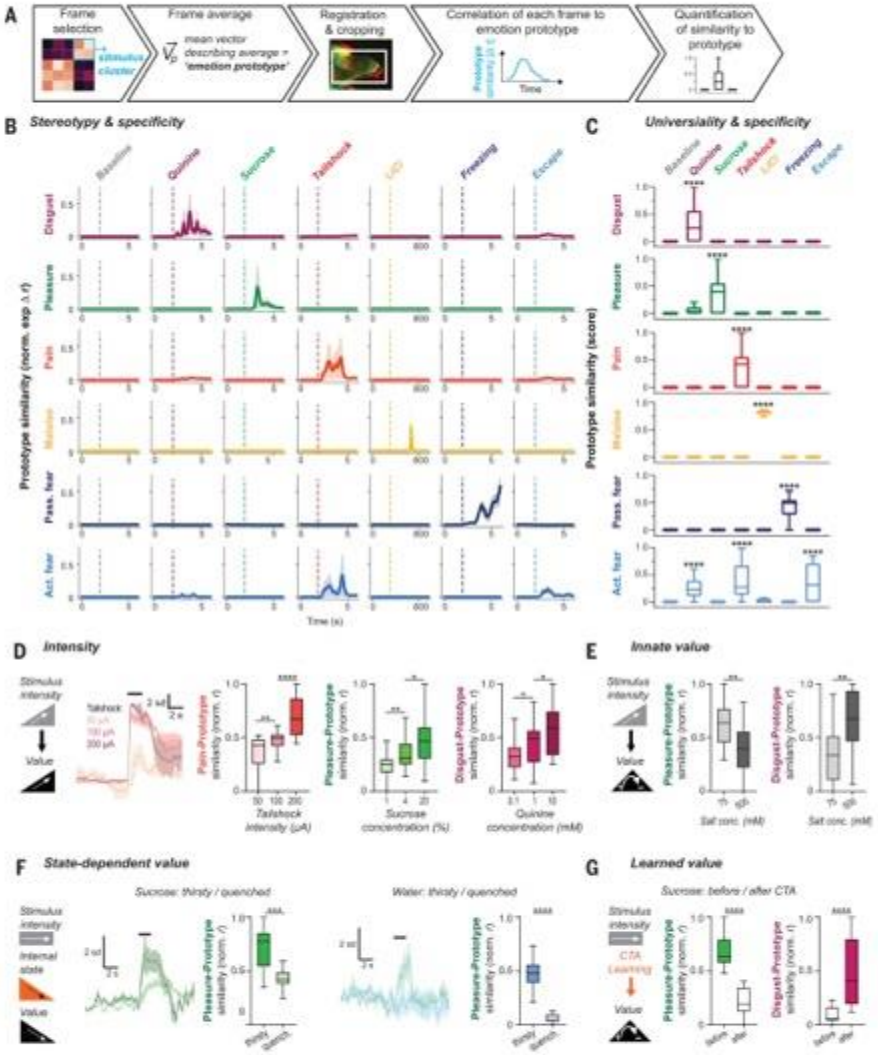


Fig. 2. Facial expressions reflect core features of emotion states. (**A**) Schematic of emotion prototype creation. (**B**) Similarities of facial expressions for each event type (three occurrences each) in one exemplary mouse to each emotion prototype. (**C**) Prototypes are valid and specific universally 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 expressions, 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

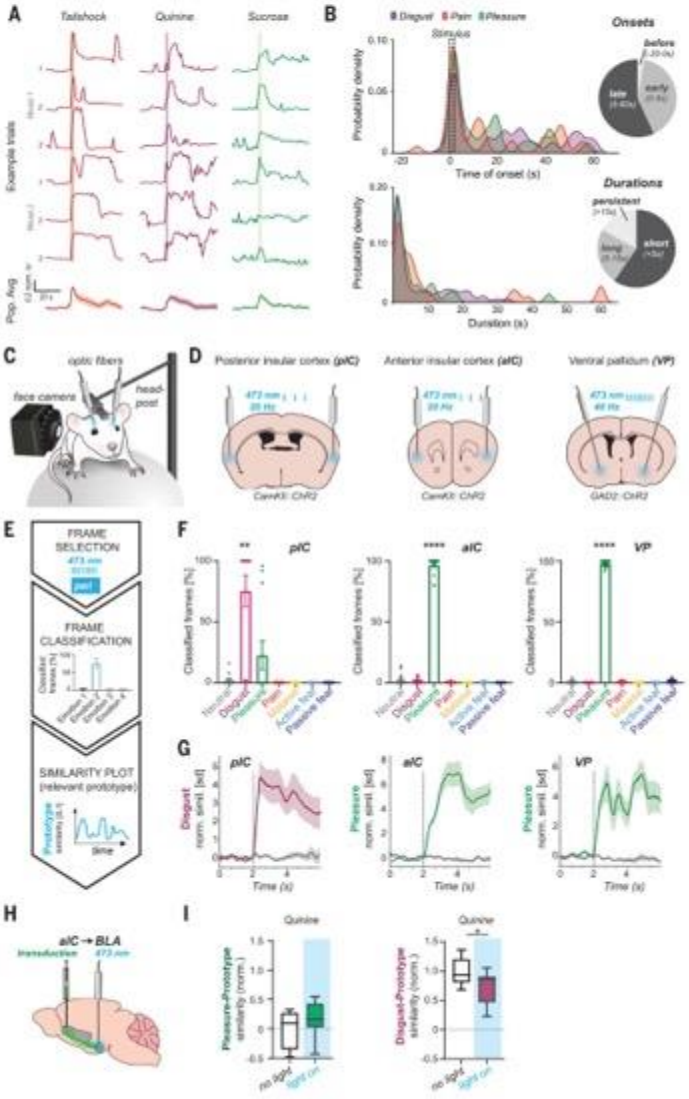


Fig. 3. Facial expressions are variable and associated with internal brain states. (**A**) Similarity traces (1-s binned) for each relevant emotion prototype (tail shock, pain prototype; quinine, disgust prototype; sucrose, pleasure prototype). (Top) Individual event-triggered facial expression traces 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.

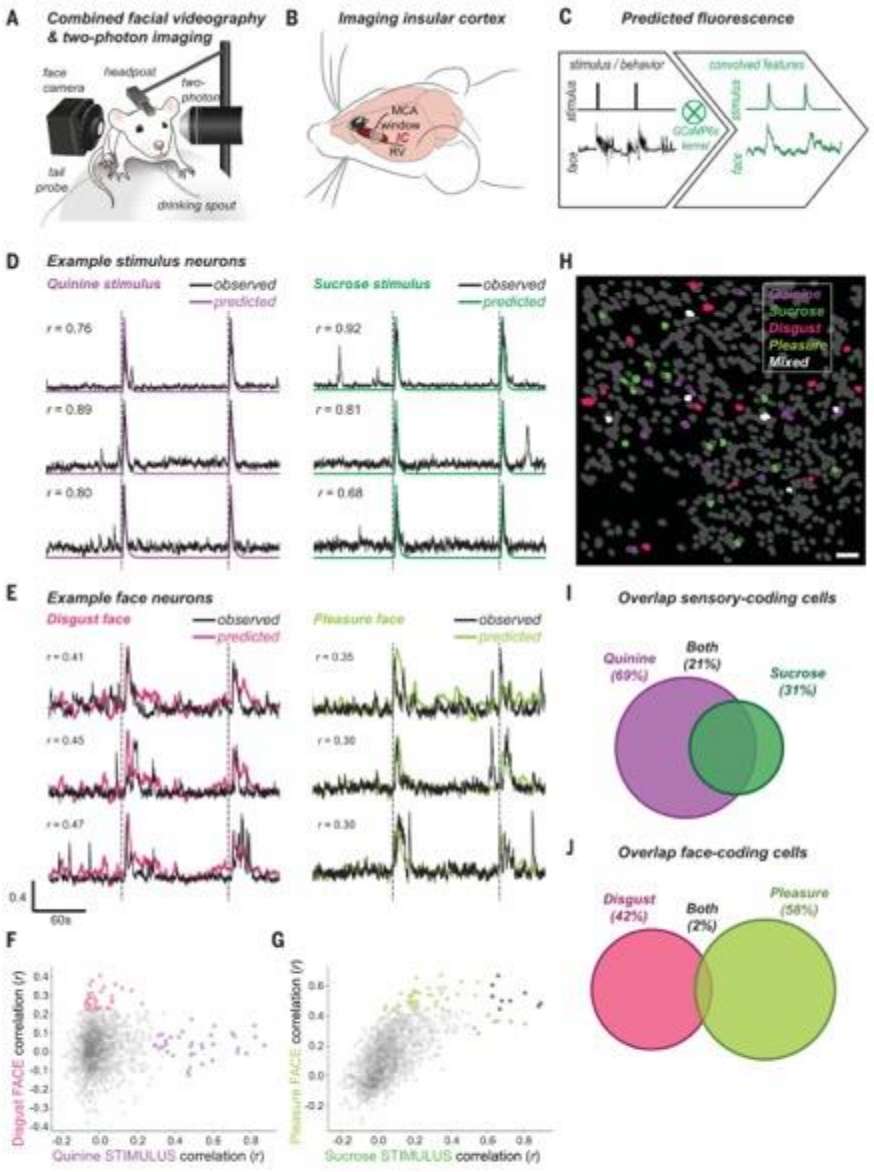


Fig. 4. Neuronal correlates of emotion state in the posterior insular cortex. (**A**) Illustration of combined facial videography with awake two-photon calcium imaging. (**B**) Schematic of the chronic window implant above the posterior insular cortex (IC, red) with respect to major blood vessels: medial 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. Wurbel, 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. Gehrlach *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. Sroka 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 (SPP1665), the German Israeli Foundation (Grant I-1301-418.13/2015), and the ANR-DFG project "SAFENET" (ANR-17-CE37-0023). **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

science.sciencemag.org/content/368/6486/89/suppl/DC1

Supplementary Materials for

Facial expressions of emotion states and their neuronal correlates in mice

Nejc Dolensek, Daniel A. Gehrlach, Alexandra S. Klein, Nadine Gogolla*

*Corresponding author: ngogolla@neuro.mpg.de

Published 3 April 2020, *Science* **368**, 89 (2020)
DOI: 10.1126/science.aaz9468

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:
(available at science.sciencemag.org/content/368/6486/89/suppl/DC1)

Movies S1 to S3

Materials and Methods

Animals

All procedures were approved by the Government of Upper Bavaria. Male and female mice at 2-6 months of age were housed in groups of 2 or 3 animals and kept on a 12 h inversed light cycle (11:00 a.m., lights off). Mice were provided with ad libitum access to standard chow and water except for drinking and CTA experiments, when mice were water deprived for periods of time specified in the following sections. All behavioral experiments were conducted with C57BL/6NRj mice, imaging experiments were performed on transgenic tetO-GCaMP6s [JAX 024742] x CaMKII-tTA [JAX 007004] mice and optogenetic experiments were performed on C57BL/6NRj and GAD2-Cre mice [JAX 010802].

Viral constructs

For in-vivo optogenetic experiments, the following constructs were obtained from the UNC Vector Core (Gene Therapy Center, University of North Carolina at Chapel Hill, USA): AAV2/5-CaMKIIa-hChR2(H134R)-eYFP ($6.2x10^{12}$ vg/ml), AAV2/5-CaMKIIa-eYFP ($4.3x10^{12}$ vg/ml), AAV2/5-EF1a-DIO-eYFP ($5.6x10^{12}$ vg/ml) and AAV2/5-EF1a-DIO-hChR2(H134R)-eYFP ($5.5x10^{12}$ vg/ml).

Surgical procedures

We injected metamizol (200 mg/kg, s.c.) for peri-operative analgesia and carprofen (s.c., 5 mg/kg, once daily for 3 days) for post-operative pain care.

Head-post implantation for facial videography

Mice were anaesthetized using midazolam / medetomidin / fentanyl mix (5mg/kg; 0.5 mg/kg; 0.05 mg/kg). Lidocaine was applied topically at the incision site. A head post was secured to the skull using C&B Super-Bond (Sun Medical).

Cranial window surgery for two-photon calcium imaging

Mice (tetO-GCaMP6s x CaMKII-tTA) were anaesthetized using midazolam / medetomidin / fentanyl mix (5 mg/kg; 0.5 mg/kg; 0.05 mg/kg) and injected with dexamethasone (s.c., 0.2 mg/kg). Lidocaine was applied topically at the incision site. A head post was secured to the skull using C&B Super-Bond (Sun Medical). Tissue including parts of the temporalis muscle was

removed and the skull was cleaned above the left insular cortex. A craniotomy was created using a 3 mm biopsy punch (Miltex) and positioned with the rhinal vein as the ventral limit and the medial cerebral artery slightly anterior to the center of the window, as seen through the skull. Since cortical surface is curved, the brain was protected from being flattened by applying a small amount of transparent polymer as previously described (*31*). A 3 mm cover slip was gently positioned on top of the polymer and secured using cyanoacrylate glue. The remaining exposed skull was covered with C&B Super-Bond.

In-vivo optogenetic experiments

Anesthesia was initiated with 5% isoflurane and maintained at 1-2.5% throughout surgery. Mice were secured in a stereotaxic frame (Stoelting), placed on a heating pad (37 °C) and eye ointment (Bepanthen, Bayer) was applied. We performed bilateral skull trepanations for all optogenetic experiments. For all viral injections, we used pulled glass-pipettes attached to a microliter syringe (5 µL Model 75 RN, Hamilton) using a glass needle compression fitting (#55750-01, Hamilton), mounted on a syringe pump controlled by a microcontroller (UMP3 + micro4, WPI). All injections were performed at a rate of 80 nl/min.

For aIC and pIC animals (C57BL/6NRj), we bilaterally injected 100 nl of virus (AAV2/5-CaMKIIa-hChR2(H134R)-eYFP or AAV-CamKIIa-eYFP for the control mice) into the aIC (distances from Bregma: AP: +1.7 mm, ML: ± 3.1 mm, DV: - 3.5 mm) or pIC (distances from Bregma: AP: -0.35 mm, ML: ± 4.05 mm, DV: - 4 mm). Custom-made optic fibers (200 µm core, 0.22 NA, 1.25 mm zirconia ferrule from Thorlabs) were implanted 0.5 mm above the injection site.

For ventral pallidum animals (GAD2-IRES-Cre) we bilaterally injected 80 nl of Cre-dependent virus (AAV2/5-EF1a-DIO-hChR2(H134R)-eYFP or AAV2/5-EF1α-DIO-eYFP for the control mice) into the ventral pallidum (distances from Bregma: AP: -0 mm, ML: ± 2.5 mm, DV: - 4.9 mm at 10° angle). Custom-made optic fibers (see above) were implanted 0.5 mm above the injection site.

For aIC→BLA projection specific manipulation, we bilaterally injected 50 nl of virus (AAV2/5-CaMKIIa-hChR2(H134R)-eYFP or AAV-CamKIIa-eYFP for the control mice) into the aIC (same coordinates as mentioned above) and bilaterally implanted optic fibers above the BLA (distances from Bregma: AP: -1.0 mm, ML: ± 3.2 mm, DV: - 3.8 mm). A head post was secured to the skull using C&B Super-Bond (Sun Medical). We sealed the space between the fiber and

the trepanation with bone wax to protect the underlying brain tissue from the cyanoacrylate glue (Ultra Gel, Pattex) that secured the optic fibers to the skull. Additionally, we added a layer of black colored dental cement to further secure the optic fibers and to reduce light emission from the skull.

Experimental timelines

All mice were handled by the experimenter for a period of at least 3 days prior to any surgical procedures. After a surgical procedure (described above), animals were allowed to rest for a period of 7-10 days. Following that, they were handled by the experimenter and habituated to head-fixation for a period of 5 days. For fixing to the head-holder on the setup, mice were briefly anesthetized with 5% isoflurane. After any head- fixation, the mice were always allowed at least 30 min to recover and habituate.

Orofacial videography

We acquired data at 30 Hz using a USB 3.0 monochrome camera (BFS-U3-13Y3M-C, Point Grey Research), positioned perpendicularly to the right side of the mouse's head. Illumination was provided by three 875 nm IR LED arrays (Kemo Electronic M120). See fig. S1 for schematic illustration of the camera and light arrangements. The camera was fitted with a NIR short-pass filter with a 900 nm cutoff (FES0900, Thorlabs) to filter out two-photon excitation light and a long-pass filter with 600 nm cutoff (FEL0600, Thorlabs) to filter out optogenetic stimulation light. Mice were weakly illuminated from behind by a 470 nm fiber coupled LED (M470F3, Thorlabs).

Stimulus deliveries and protocols

Timeline

All stimuli except where specified differently (see LiCl injection, safe/fearful context) were delivered to mice for 2 s after a baseline of 120 s in blocks of 3 repeated stimulations 120 s apart. The blocks of stimuli were separated by 500-1000 s when performed consecutively.

Stimulation sequence

Animals were exposed to sets of stimuli in blocks of 3 stimuli in a pseudo-randomized order. Strongly aversive stimuli, such as pain and malaise, were always presented as the last stimuli in an experiment.

Tastants

Tastants were presented to the animals' mouth through a round-ended reusable feeding needle (Fine Science Tools) for 2 s at a time, controlled using a solenoid pinch valve (P/N: 225PNC2-21, NResearch). Delivered tastants were: 1, 4, 20 % sucrose solution, 0.1, 1, 10 mM quinine solution, 75 and 500 mM NaCl solution (all Sigma Aldrich) and drinking water (from our animal facility). Trials were visually inspected to ensure the animals tasted the liquid presented.

Tailshocks

Weak electrical shocks (2 s; 50, 100 or 200 µA) were administered to the tail generated by a Precision animal shocker (Coulbourn Instruments). Two 0.5 mm silver-coated copper wire coils (Conrad), custom made for each mouse, were soldered to a lightweight isolated cable connected to the shocker and positioned approximately 1 cm apart at the center of the tail.

Lithium chloride induced malaise

While being head-fixed, animals were carefully injected i.p. with a lithium chloride solution (0.15 M LiCl in 0.9 % saline, 2 % body weight, e.g. 600 µl / 30 g mouse) and were recorded for a period of approximately 20 min.

Safe and fearful context

Safe context was defined as a context in which a mice has never experienced an aversive stimulus. Following a baseline recording in the safe context, mice repeatedly received unsignaled tailshocks across approximately 30 min. Mice were allowed to rest in their cage and were again placed on the behavioral setup 24 h later in what we then defined as fearful context.

Neutral / baseline condition

Neutral condition data was collected identically to all stimulus conditions, including the sampling from the same 2 s long stimulus time-points, however, no actual stimuli were presented. Furthermore, the mice were in a safe context to which they were well habituated and had no aversive experiences associated with it.

Thirst

Mice were acutely water deprived for 16-20 h and were presented with 20 % sucrose solution and water for 2 s, three times each, in a random order. Following that, they were repeatedly presented with drops of water across a period of 10 min and were allowed to drink to satiation. 5 min after that, they were again presented with drops of sucrose solution and water, identically to the start of the experiment. Trials were visually inspected to make sure that in all included trials animals tasted the liquids.

Conditioned taste aversion

Animals were acutely water deprived for 16-20 h and were presented with 20 % sucrose solution for 2 s, 3 times in the space of 10 min ("before"). Following that, they were repeatedly presented with drops of sucrose across a period of 10 min and were allowed to drink to satiation. 5 min after that the mice were carefully injected with LiCl as described above. 30 min after injection, mice were removed from the head-bar holder and returned to their cages with free access to water for 1h. After that, they were water deprived for 16-20 h, when they were head-fixed again and presented with 20 % sucrose solution for 2 s, 3 times in the span of 10 min ("after").

Optogenetic stimulation

Mice were tethered to optic patch cords and optogeneticaly stimulated with 473 nm light produced by a solid state laser (CNI Laser, China). The stimulation protocol was identical to all other stimuli, with the light being delivered for 2 s three times per experiment. The exact stimulation parameters were the following: aIC – bilateral 20 Hz stimulation, 10 ms pulse duration, 15 mW; pIC – bilateral 20 Hz stimulation, 10 ms pulse duration, 15 mW; VP: bilateral 40 Hz stimulation, 10 ms pulse duration, 15 mW; aIC→BLA, bilateral 20 Hz stimulation, 20 ms pulse duration, 10mW concurrent to 1mM quinine delivery. eYFP control mice were stimulated using the same parameters as mice expressing ChR2.

Freezing and Escape Detection

Freezing

We detected freezing episodes using a method we previously described (*22*), which establishes freezing in head-fixed animals as a combination of immobility, significant reduction in orofacial movements and a significant increase in pupil size lasting at least 2 s. We scored freezing when

either of two criteria was true: *either* an animal was immobile for at least 2 s with the z-scored pupil area was at least 1 sd higher than z-scored orofacial movement *or* a period of ≥ 5 seconds of immobility and complete lack of orofacial movements.

Pupil size measurements

We analyzed the pupil size similar to previous studies (*22*, *32*). In brief, for every frame we cropped the eye area, thresholded it to create a binary image and used the matlab function 'imfindcircles.m' to determine the pupil radius.

Orofacial movement analysis

Orofacial movements were analyzed as previously described (*33*).

Escape

Escape episodes were identified as periods of very rapid movement in a fearful context lasting at least 2 s as revealed by visual inspection.

Visualizations of typical facial expressions

Line drawings (Fig. 1B, right)

To illustrate the facial expression differences, an illustrator created line drawings of the facial outlines seen in the representative single video frames (Fig. 1B, left) selected from each emotion event.

Pixel difference heatmap overlays (Fig. 1B, right)

To visualize the areas of the face that are most changing in comparison the neutral facial expression, pixel difference heatmap overlays were created for each of the expressions in the following way: each of the example frames in Fig. 1B was pixel by pixel compared to the neutral expression frame by subtracting grayscale pixel values across the whole image. The results of this subtraction were then smoothed using a Gaussian filter with a sigma of 50 pixels, mapped to a color map, made partially transparent and overlaid over each of the original frames. Areas with almost no change were made transparent using a background removal tool for better visibility. The Gaussian filter was only applied for visualization purposes and not for data analysis.

Facial expression analyses

Motivation to employ the HOG feature approach to extract facial features of emotion

Comparing videos of behavioral features across animals is challenging because of transformations including translation, rotation, scale or changes in lighting between individuals. HOG features are designed to be largely invariant to these transformations. If the head position or illumination are slightly different between two animals, HOG will detect the pixel changes still in the same spatial grid and thus not notice these as different. Thus, the coarse binning and emphasis on orientation that HOG features provide, bear the advantage of greater invariance to changes in positioning, scale and lighting commonly occurring between different experiments. This invariance is crucial when comparing facial features across animals. Another advantage of the HOG feature extraction is the biologically inspired emphasis on edge detection. Indeed, while PCA was one of the first approaches to face recognition in humans (*34*), HOG features have been found to be superior to Eigen feature based face recognition approaches (*35*). Indeed, HOG and HOF (Histogram of optical flow) have been successfully employed in other studies of dynamic behavioral classification, especially when behavior had to be compared across individuals (*36*, *37*).

Histogram of oriented gradients descriptor creation

Histogram of oriented gradients (HOG) descriptors for each video frame were created using the following parameters: 8 histogram orientation bins, using square cells with a height of 32 pixels and 1 cell per block, with images normalized using power law compression before processing.

Similarity matrices

Face recordings of all experiments were processed using custom Python scripts. Each single frame was converted into a histogram of oriented gradients (HOG) vector. For all stimuli except LiCl injection, all frames during 4 s before the stimulus onset and 2 s during stimulus presentation, for 3 stimulus presentations, were pairwise compared (in total 540 frames). For the neutral condition, the procedure was the same, however, no actual stimulus was presented. For the LiCl injection, data was taken from a single experiment, where 3 sets of 4 s of baseline frames were selected in the first 2 minutes of experiment, after the mouse stopped exhibiting fear behaviors caused by the injection, but before malaise behavior would appear. The three 2 s long sets of stimulus frames were selected from frames 20 min after the start of experiment. A pairwise comparison of all selected frames per experiment was then performed by calculating Pearson's correlation coefficient for each possible pair and plotted in a hierarchically clustered

pairwise correlation matrix. The formed clusters were finally post-hoc compared to the frame temporal origins (before and during stimulus).

Prototypical face creation

All prototypical faces were created from a dedicated mouse whose data was not used in any other analysis. The generalizability of prototypes was tested independently and is exemplified in fig. S3A. First, a neutral prototypical facial expression was created by averaging HOGs of frames during a baseline period (250 frames) where no stimuli were presented and mouse was well habituated to the experimental setup. The validity of this baseline duration to best detect the deviations upon emotion events was independently tested and is shown in fig. S5B. The neutral prototypical face HOG was then compared to HOGs of frames during a single stimulus presentation for each of the identified states during clustering (above) by calculating a Pearson's correlation coefficient for each. For each of the states, 10 frames most dissimilar to the neutral prototypical face (the ones with the lowest correlation coefficient) were selected and averaged, resulting in a single HOG for each respective emotion. Prototypical faces were specifically build based on data associated with the following stimuli: disgust – 10mM quinine solution; pleasure – 20 % sucrose solution; malaise - 0.15 M LiCl i.p. injection; pain – 200 µA tail shock; passive fear - single detected freezing episode in a fearful context; active fear – single detected escape episode in fearful context. We independently confirmed that building the prototypes on medium stimulus intensities would still reveal scalable facial expressions. We found, however, that building the prototype on the strongest emotion event yielded the best prototype. The scalability was lost when prototypes were built on too low stimulus intensities (shown for pain in fig. S5C).

Mouse face registration and cropping

All data acquired using facial videography was registered to a single frame of a wild-type male mouse head-fixed to the setup exhibiting a neutral facial expression ("template frame"). This was achieved using custom written Python code employing a phase correlation approach, which enables estimation of relative offset of two similar images. For each recording, a single frame at the start of experiment was checked for optimal registration parameters, which were then applied across the whole recording. The image was first roughly aligned by being moved in relation to the template frame in the XY axis to the position of lowest offset, as indicated by phase correlation analysis. Following that, the image was angle aligned by being rotated for a single

degree at a time 45 degrees upwards and 45 degrees downwards, with the angle of lowest offset to the template frame being selected and applied. After correcting the rotation, the image was once again aligned in the XY axis. Finally, the image was scale matched to the template by rescaling it between 0.01 and 2 times in 0.01 steps and identifying the scale of lowest error based on phase correlation results, which was applied to the image as the final step. After the registration procedure was finished, the image was cropped using the same set of coordinates manually created to crop the template frame. The cropping coordinates were carefully selected to contain as little of non-face areas as possible, while not cutting off any of the face at a baseline facial expression, but also during any of the facial behaviors like licking. Attention was paid to also contain a section of empty space directly in front of the face in the template frame, since whiskers will move into that area during expressions such as pain and escape.

Similarity to prototypical face

To measure similarity to various prototypical faces, each acquired frame from any of the experiments was registered, cropped and converted into a HOG descriptor vector, as described above. Following that, each frame ("target frame") can be compared to a chosen prototypical face for similarity by calculating Pearson's r between the target frame HOG vector and the prototypical face HOG vector. If this is done across a single continuous experiment, the similarity comparison will result in a single similarity value per each time point, which can be plotted in a sequence, revealing a temporally evolving intensity graded description of experience.

Facial expression onsets and duration

We detected facial expression onsets (as in Fig. 3B, top) when the mean prototype similarity in a 1s bin was at least 2sd above baseline. The durations (as in Fig. 3B, bottom) were calculated from the number of consecutive 1s bin which were at least 2sd above baseline prototype similarity.

Classifiers

To confirm quantitative distinctiveness between facial expressions evoked by different stimuli or behaviors and create a tool that is able to separate any given set of frames into different discrete emotions, we created and trained a random forest classifier. The training dataset consisted of data collected from 3 animals, with 3 stimulus presentations for each emotion per each animal. We selected all 30 frames from the 2nd second of each stimulus presentation and appropriately

manually labeled them based on experimental origin. All frames were registered, cropped and converted into HOGs. Prior to classifier training, all HOG features corresponding to the area that could contain the spout (in all conditions whether spout was present or not), corresponding to the bottom right corner, size 4x9 HOG cells, were set to 0 to prevent any influence of the spout presence on prediction. To test the performance of the classifier, the training dataset was split into a smaller training dataset containing a randomly selected sample of 15% of all training data and tested on the remaining 85%. This was repeated 1000x, with the performance evaluated at each repetition. To establish a baseline performance to compare to, the classifier was trained and repeatedly tested in an exactly same manner, however, prior to training, all training dataset labels were randomly shuffled, while the performance was evaluated on the correctly labelled data.

Visualization of classifier feature importance (fig. S4)

To visualize importance of different sections of the face for decision 7 binary random forest classifiers were trained, one for each emotion state described in this manuscript, with all frames manually labeled as "1" (belonging to the emotion detected by the classifier) or "0" (not belonging to the specific emotion detected by the classifier). The data used for training was same as for the multiclass classifier described above, with the exception that no splitting into test and training datasets was performed. Since each HOG cell contains multiple orientation bins which cannot be easily visualized, the overall importance of each section of the frame was calculated as the mean feature importance of the 8 underlying orientation bins, for each classifier.

2-dimensional embedding of facial expressions

All video frames during the 2nd second of each stimulus presentation (30 frames) were collected for three stimulus repetitions for each respective stimulus/state. These frames were registered, cropped and converted into HOG vectors (5040 dimensions) and then dimensionality reduced using principal component analysis (100 dimensions). Following that, t-SNE (2 dimensions) was ran on the data of each respective mouse. The 2 t-SNE dimensions describing each data point were used to create a scatter plot, with each point representing a single frame. These points were finally post-hoc labeled based on the type of stimulus presented during their respective origins.

Optogenetic facial videography pipeline

To analyze the data collected in the optogenetic stimulation experiments, all frames at times of interest were first registered to the template frame, cropped and converted into HOG descriptors,

as described before (three 2s optogenetic stimulations per condition per animal). Frames belonging to a set of 3 stimulations from a single animal were then pairwise similarity compared by calculating Pearson's r for each pair and hierarchically clustered to reveal distinct groups of facial expressions corresponding to the peri-stimulus epochs. Following that, frames from the peri-stimulus epochs were processed using a random forest classifier trained on wild-type mice as described in the *classifier* section above. This resulted in a distribution of discrete emotion labels which guided selection of an appropriate prototypical face (based on the emotion most represented in the classifier detections). This prototypical face was then used as described above in the *Similarity to prototypical face section* to produce a graded readout of experience of the detected emotion.

Awake head-fixed two-photon imaging

Image acquisition

Two-photon imaging was carried out using a rotating resonant-scanning two-photon microscope (B-scope, Thorlabs) set to an angle perpendicular to the insular cortex surface (75 – 85° rotation to access the left hemisphere) with a 16x water immersion objective (Nikon N16XLWD-PF, N.A. = 0.8). This provided an 830 x 830 µm field of view that was scanned at 14.8 Hz with a resolution of 512 x 512 pixels. All imaged fields of view were imaged at a depth of 150-250 µm below the brain surface, using a Mai Tai DeepSee laser (Newport Corp.) set to 940 nm and a power of 12-30 mW at the front aperture of the objective.

Two-photon data pre-processing

Two-photon acquired neural data was processed using the two photon analysis toolbox Suite2p: https://github.com/MouseLand/suite2p (*38*), which was used to perform motion correction, neuron detection, cell-segmentation and fluorescence measurement over time for each cell. For each field of view, the cell segmentation was manually reviewed and corrected when necessary.

Neural data analyses

To quantify the degree of relatedness of each neuron's activity to either the facial expression or stimulus, we employed a previously described (*39*, *40*) convolution approach to create predicted fluorescence traces for each (stimulus, facial expression) by convolving them with a GCaMP6s kernel. More specifically, we represented the stimulus as a square wave set to 1 when the stimulus was being presented (3 times 2s) and 0 everywhere else with a sampling rate of 30 Hz

and we used a relevant prototypical face similarity trace to represent the concurrent facial expression. Both of these were then convolved with a GCaMP6s kernel to produce a convolved stimulus trace and a convolved facial expression trace, which we compared to fluorescence of single neurons. Extracted fluorescence traces for each neuron were upsampled to facial videography sampling rate (30Hz), to enable comparison to both of the convolved traces, which we performed by calculating Pearson's r. We correlated each neuron to the relevant convolved stimulus traces and convolved facial expression trace. To determine which neurons were significantly correlated to either the face or the stimulus, we followed the approach described in (*40*). Specifically, we determined the thresholds for a significant correlation in an unbiased manner by calculating a null distribution for each imaging experiment by splitting all face/stimulus traces into 11 equally sized sections, randomly shuffled their temporal order, convolved the resulting traces and correlated them to fluorescence of all concurrently recorded neurons. This was repeated 1000 times for each dataset and the 99th percentile of the resulting null distribution was chosen as the significant correlation threshold.

Human Raters Experiment

To evaluate whether untrained human observers would be able to detect emotional expression in mice, volunteers at a science outreach event answered a computer based questionnaire. Sixty-three human volunteers were shown single image examples of mice exhibiting neutral, passive fear, pleasurable, disgusted or painful facial expressions and asked two questions. 1) "Is this expression good or bad?" (Options to select good or bad) and 2) "What is the underlying experience?" (Options to select neutral, pleasure, disgust, fear and pain). We did not control for participant's age, sex or previous experience in working with mice.

Statistical Analysis

Analyses were performed using either Graphpad Prism (GraphPad Software, Inc., La Jolla, CA, USA, Version 8) or Python. Group comparisons were made using one-way ANOVA followed by Dunnett's post hoc tests comparing all groups to a baseline group in case main effect of ANOVA was statistically significant ($p < 0.05$). Two-group single variable comparisons were performed using a two-tailed Mann-Whitney U test. All correlation analyses were performed by calculating Pearson's correlation coefficient. To compare groups to a predetermined value, one sample two-sided Wilcoxon test was used. Detailed information about the type and results of all statistical

procedures can be found in Table S1. All animal numbers are reported in Figures and their legends. No statistical methods were used to predetermine sample size. Since we did not assume our data to be normally distributed we used nonparametric statistics except for the group comparisons in Fig. 2.

Supplementary Figures

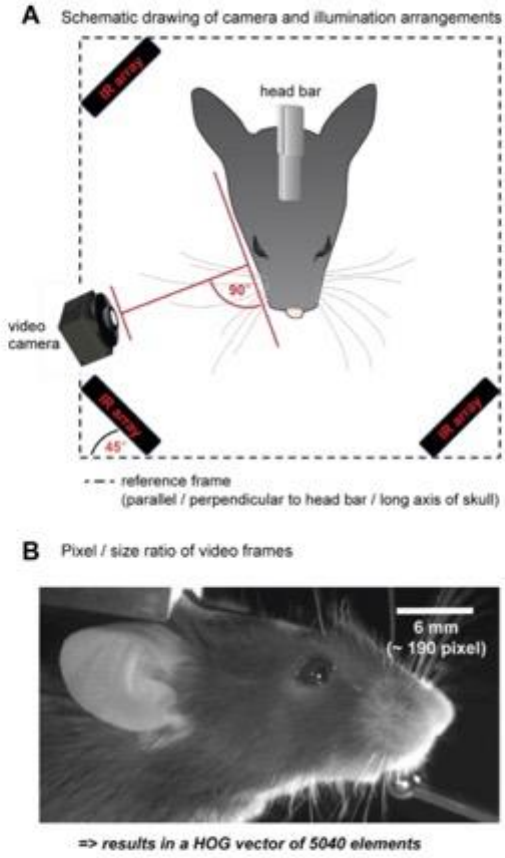


Fig. S1. Facial videography setup

(**A**) Illustration of camera and illumination arrangement with respect to the head of the mouse and the head bar. Three infrared arrays were used and mounted to a frame constructed around the head-fixation set-up. The frame was aligned to the head-bar and long-axis of the mouse skull as illustrated and infrared arrays were positioned at 45° angles to this frame. The camera was positioned at 90° to a virtual line along the mouse's lateral face surface. Camera positioning by hand was enough to ensure optimal video acquisition. We estimate that deviations of up to 15° from the ideal angle may still be tolerated as long as all facial features are in focus (ears, nose, eyes). The distances of camera and illumination to the head depended on the camera lenses used. Therefore, we illustrated in (**B**) the video frame resolution (cropped image that was used to create the HOG) with reference to the head size. We used a field of view of 1144 x 586 pixels, with a pixel resolution of approximately 30 pixels / mm. The mouse face registration and cropping procedure described in the methods section was employed to realign mouse faces and make them comparable, whenever camera angle and or distance varied between experiments.

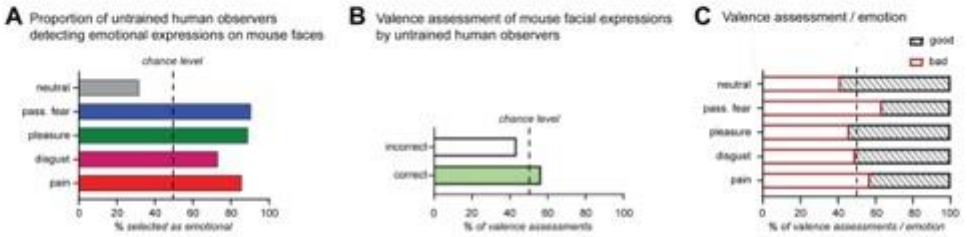


Fig. S2. Untrained human observers readily detect emotional facial expressions in mice.

Sixty-three human volunteers were asked in the context of a science outreach event to rate mouse facial expressions without further instructions. They were given two questions: 1) 'Is this expression good or bad?' and 2) 'What is the underlying experience?' (multiple choice format). (**A**) Untrained human observers reliably detected emotional expressions above chance and correctly rated neutral expressions as non-emotional. (**B**) The same observers failed to assign the correct valence (good or bad) to the emotional expressions. (**C**) However, a trend towards correctly classifying pleasure as good, but passive fear (freezing) and pain as bad was observed. Interestingly, also neutral expressions were more frequently rated as 'good'.

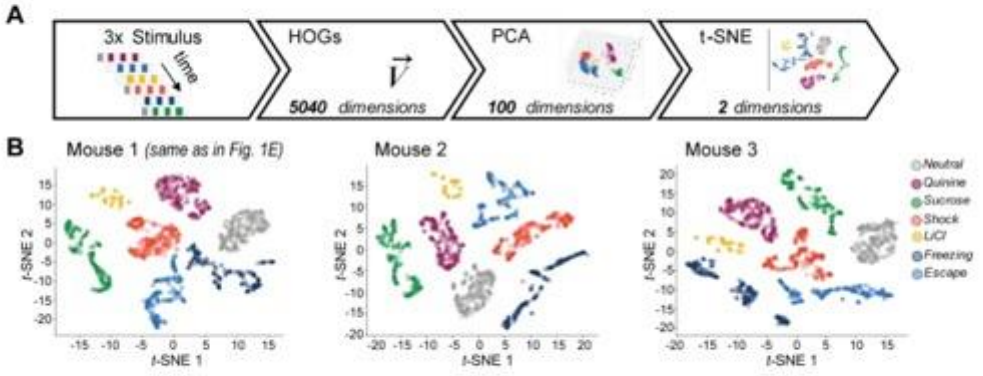


Fig. S3. t-SNE visualization in three individual mice.

(**A**) Analysis pipeline for t-SNE visualizations (**B**) t-SNE visualizations in each individual demonstrates clean separation into event-specific clusters.

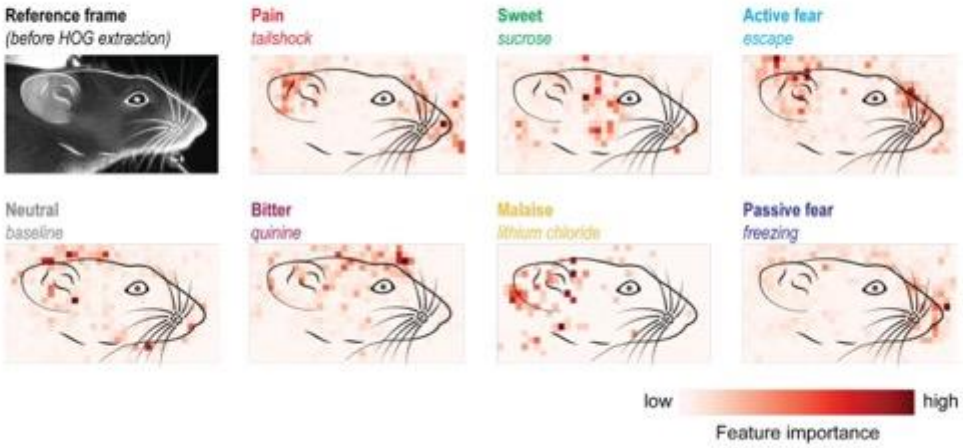


Fig. S4. Illustration of feature importance used by the random forest classifier to distinguish facial expressions of different emotion categories

The feature importance of HOG vector variables that were most decisive for the random forest classifier to distinguish different facial expressions were determined and the corresponding areas were color coded in an illustration. The color code depicts the level of feature importance to take a decision. A line drawing derived from the neutral expression is overlaid on all events to facilitate orientation.

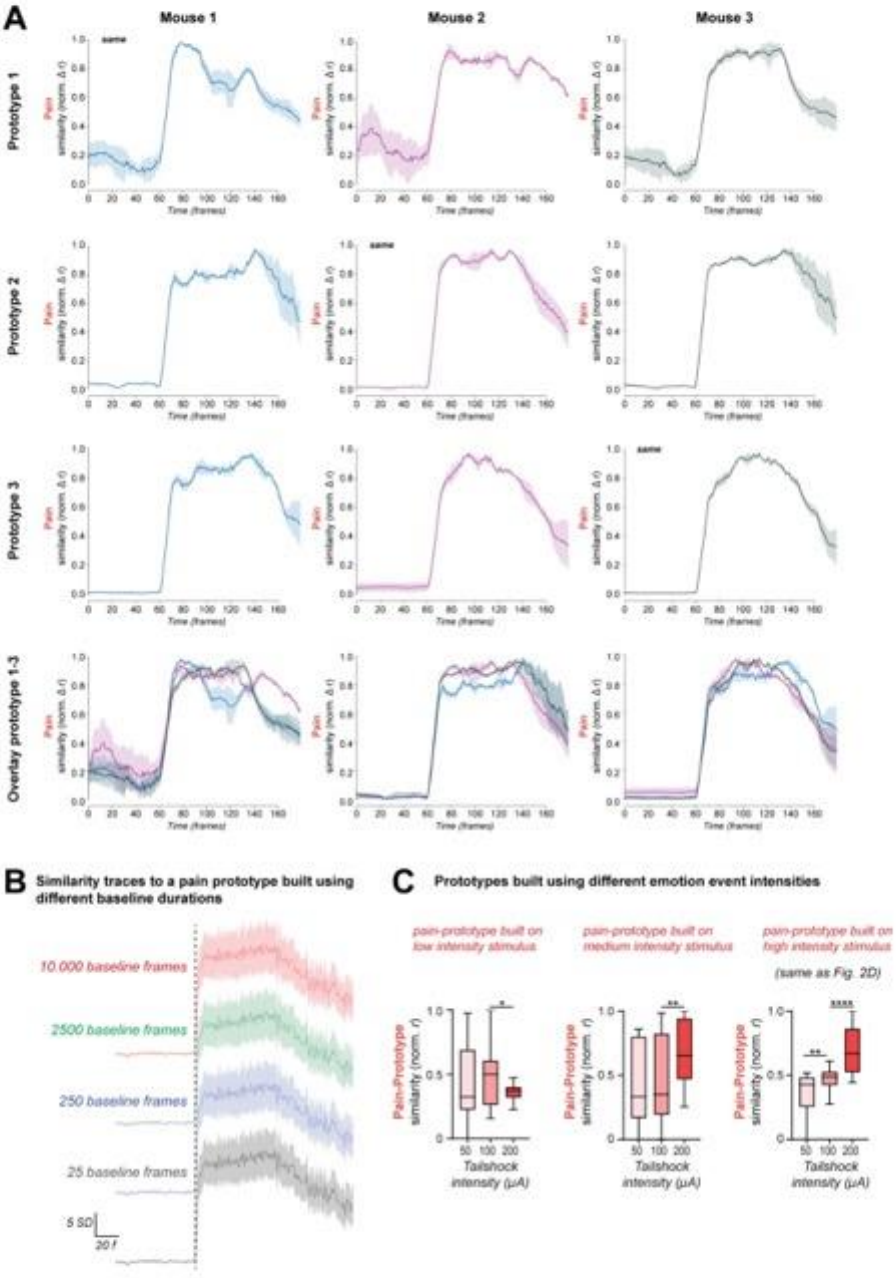

A
Mouse 1
Mouse 2
Mouse 3
Prototype 1
Prototype 2
Prototype 3
Overlay prototype 1-3
same
Pain
similarity (norm. Δ r)
Time (frames)
B
Similarity traces to a pain prototype built using different baseline durations
10.000 baseline frames
2500 baseline frames
250 baseline frames
25 baseline frames
5 SD
20 f
C
Prototypes built using different emotion event intensities
pain-prototype built on low intensity stimulus
pain-prototype built on medium intensity stimulus
pain-prototype built on high intensity stimulus
(same as Fig. 2D)
Pain-Prototype similarity (norm. r)
Tailshock intensity (µA)

Fig. S5. Universality of emotion-prototypes.

(**A**) Prototypes can be built on any mouse and reveal stereotyped facial expressions in any other mouse. Here, as an example, we built one pain prototype on three individual mice and tested the validity of the prototype on each of them. All prototypes clearly reveal painful expressions in each of the mice. Notably, the prototype built and tested within the same mouse (marked as 'same') does not perform any better than a prototype built on a different mouse. Stimulus onset was at 60 frames = 2 s.

(**B**) Prototypes are built from frames the most different from a baseline ('neutral prototype'). Neutral prototypes are efficient in characterizing the emotion feature frames whether they are built from very few frames (25) or very many (2500, or 10.000). We here chose to build the neutral condition on 250 baseline frames.

(**C**) Prototypes can be built from differently strong emotion events. Comparing the scalability of emotions detected by similarities to prototypes on low, medium or strong stimulus intensities revealed that the best results are achieved when the prototype is built on the strongest emotion event. While prototypes built on medium intensity events still reveal scalability to some extent, this feature is lost when the prototype is built on the lowest stimulus intensity.

Two-tailed Mann-Whitney test: low intensity prototype (left), 50 vs. 100 µA: U = 288, P = 0.1902 and 100 vs 200 µA: U = 223, *P = 0.0137; medium intensity prototype (middle), 50 vs. 100 µA: U = 335, P = 0.6186 and 100 vs 200 µA: U = 188, **P = 0.0019; highest intensity prototype (right), 50 vs. 100 µA: U = 189, **P = 0.002 and 100 vs 200 µA: U = 127, ****P < 0.0001.

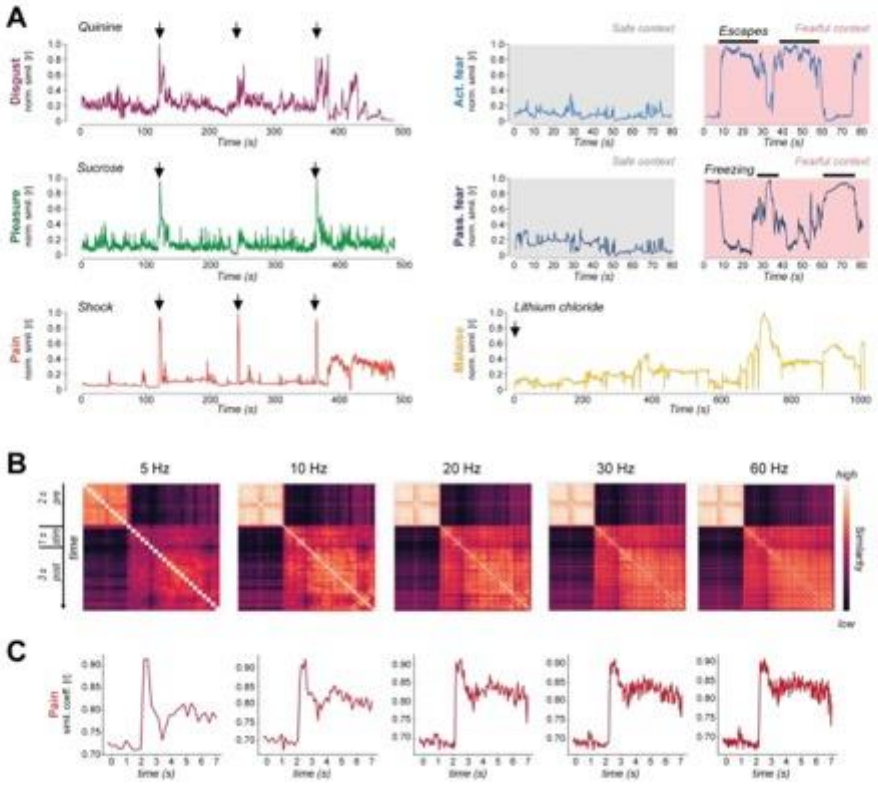


Fig. S6. Temporally resolved descriptions of facial expressions in mice.

(**A**) Exemplary similarity traces from individual animals to the prototypical expressions. Note that the face comparisons reveal diverse temporal dynamics of facial expressions in reaction to distinct stimuli / behavior onsets. All traces are from independent animals, except active and passive fear, which are the same expression compared to active and passive fear prototypes in a temporally aligned manner and while exposed to safe or fearful environments. Note the alternating pattern of freezing and escapes, which is reliably detected by facial tracking. Arrows denote stimulus onset. Horizontal black bars represent behavioral expressions.

(**B**) Comparison of different sampling frequencies (5-60 Hz) from one mouse experiencing a tailshock. The shock is administered for 1 second after 2 seconds. All frames were transformed into HOG vectors and pairwise similarity matrices were plotted. The HOG are plotted in temporal sequence of acquisition. Note that at all imaging frequencies the frames before the tailshock exhibit high similarity and are clearly distinct from the frames after tailshock.

(**C**) Similarity traces to the pain prototype built on all frames included within 400 ms during the tailshock at the different sampling rates. The shock is administered for 1 second after 2 seconds. Note that the similarity to the prototype is clearly detectable at all sampling frequencies. Higher sampling rates of 20-30 Hz reveal a more detailed description of the facial dynamics than a sampling rate of 5 Hz, however acquisition at 60 Hz does not qualitatively improve the facial description over 30 Hz.

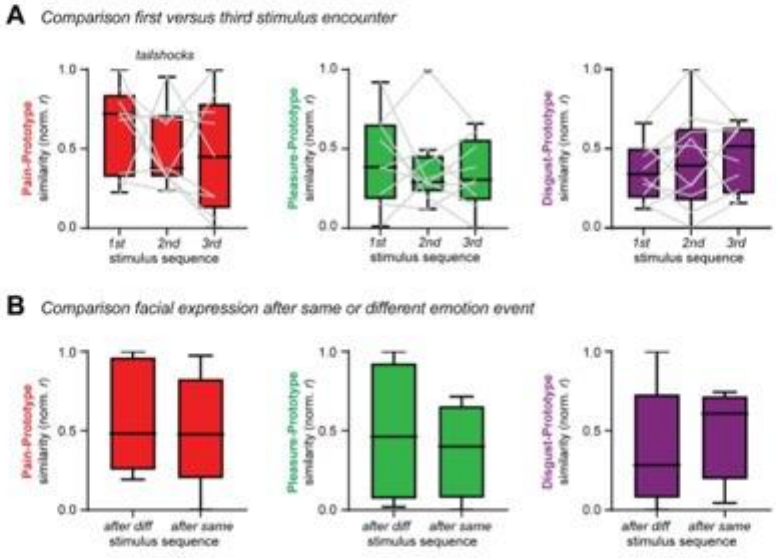


Fig. S7. Influence of stimulation sequence on emotional facial expressions (at 120 second ISI).

(**A**) Comparison of emotional facial expressions exhibited upon the first, second or third encounter of a given emotion event. The first encounter with a painful (tailshock), pleasurable (sucrose) or disgusting (quinine) stimulus does not evoke different facial expressions when compared to the third encounter. Box-and-whisker plots are Tukey style. Overlaid on the box-and-whisker plot are the changes from 1st to 3rd encounter in individual mice. N= 3 mice, n = 9 trials. Two-tailed RM one-way ANOVA: Left, pain: P = 0.3946, F (1.834, 14.67) = 0.9711; middle, pleasure: P = 0.6290, F (1.582, 12.66) = 0.4047; right, disgust: P = 0.4140, F (1.728, 15.55) = 0.8965.
(**B**) Comparison of emotional facial expressions following the same or a different emotion event. The facial expressions following the same type of emotion event do not consistently differ from the ones which follow the same type of emotion event. Box-and-whisker plots are Tukey style. Two-tailed Mann-Whitney U test: Left, pain: P = 0.8413, U = 11; middle, pleasure: P = 0.7302, U = 8; right, disgust: P = 0.5476, U = 9.

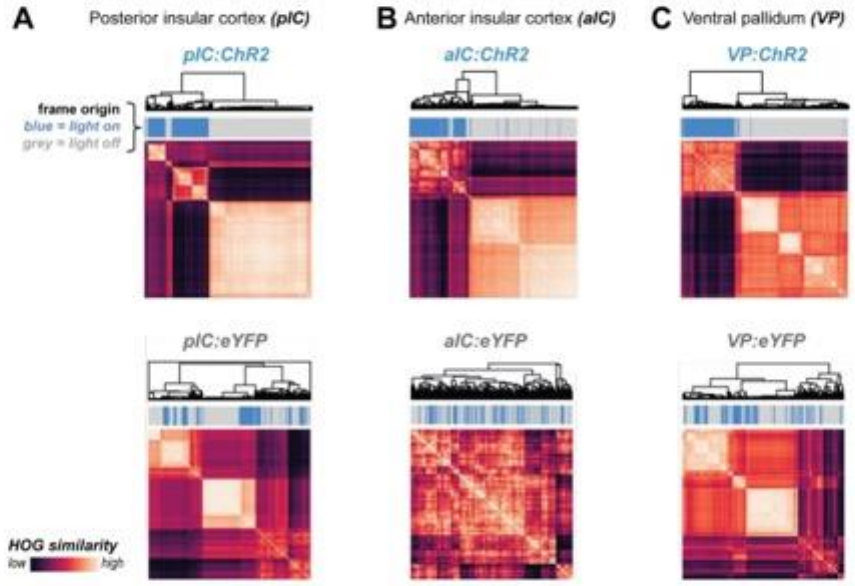
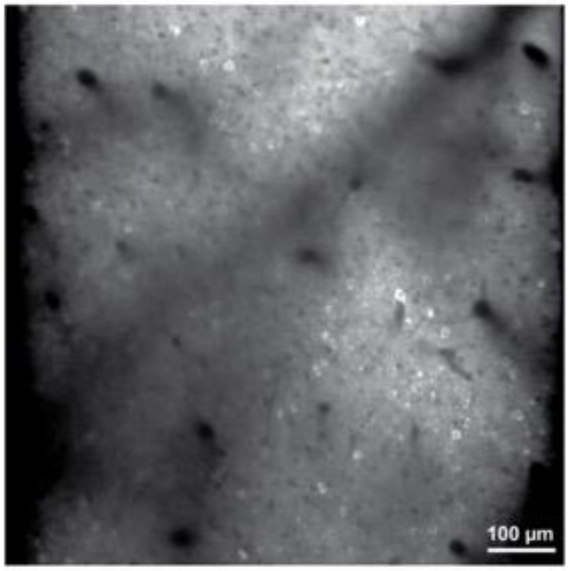


Fig. S8. Similarity matrices from individual animals exposed to optogenetic circuit manipulations.

(**A-C**) Top row: examples pairwise correlations of all frames before and during blue light stimulation for animals expressing channelrhodopsin. Bottom row: examples of pairwise correlations of all frames before and during blue light stimulation for animals expressing eYFP as control protein. Each mouse was stimulated three times for 2 second. Matrices are clustered in an unsupervised manner by similarity. Above the matrices are post-hoc temporal assignments with frames before the light in grey and frames during light in blue. Above the frame annotation are dendrograms revealing the clusters. Note the strong clustering of HOG vectors corresponding to the light *on* frames (blue) in channelrhodopsin-expressing animals which is absent in eYFP-expressing mice. The same is also revealed in the dendrograms on the top.

A *Examplary two-photon image*



B *HOG features in mice with imaging window implant above insular cortex*

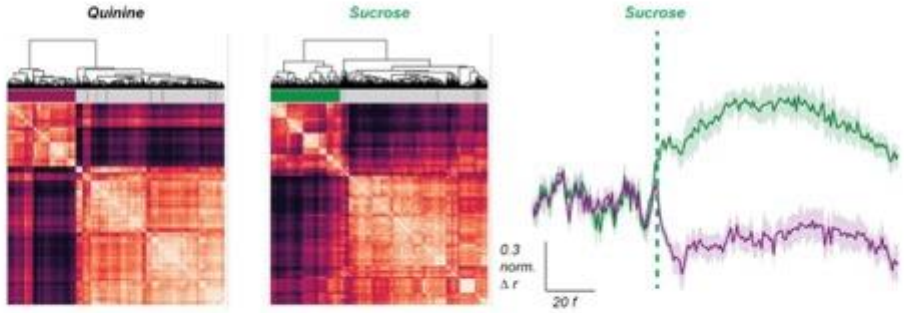


Fig. S9. Simultaneous two-photon imaging in the insular cortex and facial expression analyses.

(**A**) The same imaging field as shown in Fig. 4H. Image is a standard deviation projection of 7500 frames (i.e. ~500 s) imaged at 14.8 Hz. During this period, the animal was exposed three times for two seconds to a sucrose solution. The field of view is 830 x 830 µm large (resolution: 512 x 512 pixels).
(**B**) HOG-features before and during quinine- (left) or sucrose (middle) exposure cluster together in mice which had cranial-window implants above the left insular cortex, while facial features were analyzed from the right lateral facial surface. Right: Similarity to prototypical pleasure (green) or disgust (purple) face upon 3 exposures to 2-seconds of sucrose. HOG-extracted facial features are thus not affected by the implantation of a cranial window on the opposite side of the head. Shaded areas are s.e.m..

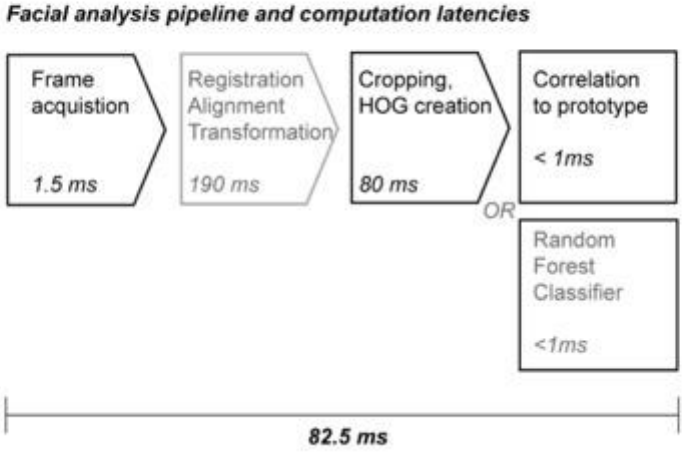


Fig. S10. Timeline of computation latencies.
Illustration of steps and latencies involved to acquire, process and compare/classify a live recorded video frame to a facial expression category. The registration, alignment and transformation step is only required when images were transformed between imaging sessions. When prototypes are build (or the classifier is trained) on animals from the same cohort and camera position, this step can be omitted. The total computation pipeline (total ~82.5 ms) can be run at 10 Hz and thus may allow for closed-loop configurations.

Table S1. Detailed statistics

Fig. 1F	Random forest classifier	neutral vs. shuffled neutral	Mann-Whitney test	two-tailed	U = 0	****P<0.0001
		disgust vs. shuffled disgust	Mann-Whitney test	two-tailed	U = 0	****P<0.0001
		pleasure vs. shuffled pleasure	Mann-Whitney test	two-tailed	U = 0	****P<0.0001
		malaise vs. shuffled malaise	Mann-Whitney test	two-tailed	U = 0	****P<0.0001
		pain vs. shuffled malaise	Mann-Whitney test	two-tailed	U = 0	****P<0.0001
		passive fear vs. shuffled passive fear	Mann-Whitney test	two-tailed	U = 0	****P<0.0001
		active fear vs. shuffled active fear	Mann-Whitney test	two-tailed	U = 0	****P<0.0001

Fig. 2C	Disgust prototype	Ordinary one-way ANOVA F (6, 182) = 20.53 ****P<0.0001	Quinine vs. neutral	Dunnett's	****	<0.0001
			Sucrose vs. neutral	Dunnett's	n.s.	0.9996
			LiCl vs. neutral	Dunnett's	n.s.	>0.9999
			Tailshock vs. neutral	Dunnett's	n.s.	0.8527
			Freezing vs. neutral	Dunnett's	n.s.	>0.9999
			Escape vs. neutral	Dunnett's	n.s.	>0.9999
	Pleasure prototype	Ordinary one-way ANOVA F (6, 182) = 33.42 ****P<0.0001	Quinine vs. neutral	Dunnett's	n.s.	0.3197
			Sucrose vs. neutral	Dunnett's	****	<0.0001
			LiCl vs. neutral	Dunnett's	n.s.	>0.9999
			Tailshock vs. neutral	Dunnett's	n.s.	>0.9999
			Freezing vs. neutral	Dunnett's	n.s.	>0.9999
			Escape vs. neutral	Dunnett's	n.s.	>0.9999
	Pain prototype	Ordinary one-way ANOVA F (6, 182) = 20.09 ****P<0.0001	Quinine vs. neutral	Dunnett's	n.s.	0.8267
			Sucrose vs. neutral	Dunnett's	n.s.	>0.9999
			LiCl vs. neutral	Dunnett's	n.s.	>0.9999
			Tailshock vs. neutral	Dunnett's	****	<0.0001
			Freezing vs. neutral	Dunnett's	n.s.	0.9954
			Escape vs. neutral	Dunnett's	n.s.	>0.9999
	Active fear prototype	Ordinary one-way ANOVA F (6, 182) = 22.81 ****P<0.0001	Quinine vs. neutral	Dunnett's	****	<0.0001
			Sucrose vs. neutral	Dunnett's	n.s.	0.9562
			LiCl vs. neutral	Dunnett's	n.s.	0.9562
			Tailshock vs. neutral	Dunnett's	****	<0.0001
			Freezing vs. neutral	Dunnett's	n.s.	>0.9999
			Escape vs. neutral	Dunnett's	****	<0.0001
	Pass. fear prototype	Ordinary one-way ANOVA F (6, 182) = 45.08 ****P<0.0001	Quinine vs. neutral	Dunnett's	n.s.	>0.9999
			Sucrose vs. neutral	Dunnett's	n.s.	>0.9999
			LiCl vs. neutral	Dunnett's	n.s.	>0.9999
			Tailshock vs. neutral	Dunnett's	n.s.	0.1420
			Freezing vs. neutral	Dunnett's	****	<0.0001
			Escape vs. neutral	Dunnett's	n.s.	>0.9999
	Malaise prototype	Ordinary one-way ANOVA F (6, 182) = 2428 ****P<0.0001	Quinine vs. neutral	Dunnett's	n.s.	>0.9999
			Sucrose vs. neutral	Dunnett's	n.s.	>0.9999
			LiCl vs. neutral	Dunnett's	****	<0.0001
			Tailshock vs. neutral	Dunnett's	n.s.	>0.9999
			Freezing vs. neutral	Dunnett's	n.s.	>0.9999
			Escape vs. neutral	Dunnett's	n.s.	>0.9999

Fig. 2D	tailshock strength	50 vs. 100 µA	Mann-Whitney test	two-tailed	U = 189	**P = 0.002
	tailshock strength	100 vs. 200 µA	Mann-Whitney test	two-tailed	U = 127	****P < 0.0001
	sucrose concentration	1% vs. 4 %	Mann-Whitney test	two-tailed	U = 192	**P = 0.0024
	sucrose concentration	4% vs. 20%	Mann-Whitney test	two-tailed	U = 236	*P = 0.0257
	quinine concentration	0.1mM vs. 1 mM	Mann-Whitney test	two-tailed	U = 294	*P=0.0207
	quinine concentration	1mM vs. 10mM	Mann-Whitney test	two-tailed	U = 291	*P=0.0181
Fig. 2E	salt – pleasure similarity	75 vs. 500mM	Mann-Whitney test	two-tailed	U = 43	**P=0.0031
	salt – disgust similarity	75 vs. 500mM	Mann-Whitney test	two-tailed	U = 47	**P=0.0055
Fig. 2F	sucrose – pleasure similarity	thirst vs. quenched	Mann-Whitney test	two-tailed	U = 33	***P=0.0006

	water – pleasure similarity	thirst vs. quenched	Mann-Whitney test	two-tailed	U = 10	****P<0.0001
Fig. 2G	pleasure similarity	sucrose before vs. after CTA learning	Mann-Whitney test	two-tailed	U = 0	****P<0.0001
	disgust similarity	sucrose before vs. after CTA learning	Mann-Whitney test	two-tailed	U = 12	****P<0.0001

Fig. 3F	Posterior insular cortex stimulation	One sample Wilcoxon test (sample: 0.1430)	Discrepancy: Neutral: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Disgust: 0.8570	W = 66	p = 0.0039
			Discrepancy: Pleasure: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Pain: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Malaise: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Act. Fear: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Pass. Fear: -0.1430	Median and mean below sample	Median and mean below sample
Fig. 3F	Anterior insular cortex stimulation	One sample Wilcoxon test (sample: 0.1430)	Discrepancy: Neutral: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Disgust: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Pleasure: 0.8570	W = 171	p < 0.0001
			Discrepancy: Pain: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Malaise: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Act. Fear: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Pass. Fear: -0.1430	Median and mean below sample	Median and mean below sample
Fig. 3F	Ventral pallidum stimulation	One sample Wilcoxon test (sample: 0.1430)	Discrepancy: Neutral: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Disgust: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Pleasure: 0.8370	W = 171	p < 0.0001
			Discrepancy: Pain: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Malaise: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Act. Fear: -0.1430	Median and mean below sample	Median and mean below sample
			Discrepancy: Pass. Fear: -0.1430	Median and mean below sample	Median and mean below sample

Fig. 3I	aIC→BLA stimulation	Pleasure no light versus with light	Mann-Whitney test	two-tailed	U = 28	P = 0.2973
		Disgust no light versus with light	Mann-Whitney test	two-tailed	U = 17	*P = 0.0400

Captions for Supplementary Movies

Movie S1

Facial videograph of a head-fixed mouse drinking sweet sucrose solution (left) or bitter quinine solution (right). Note the striking difference in facial expressions upon exposure to this appetitive or aversive experiences.

Movie S2

Simultaneous playback of the facial videograph of a facial expression of pain in reaction to a tailshock and a similarity plot (Pearson's r) to the prototypic pain expression built on a different mouse. Note the high specificity to the pain reaction.

Movie S3

Side-by-side comparison of the facial expression within the same mouse exposed to a sweet sucrose solution (left) or optogenetic stimulation of GABAergic neurons in the ventral pallidum (right). Note the striking similarity of the evoked facial expression.

References and Notes

1. C. Darwin, *The Expression of the Emotions in Man and Animals* (London, Murray, 1872).

2. A. Damasio, G. B. Carvalho, The nature of feelings: Evolutionary and neurobiological origins. *Nat. Rev. Neurosci.* **14**, 143–152 (2013). doi:10.1038/nrn3403 Medline

3. D. J. Anderson, R. Adolphs, A framework for studying emotions across species. *Cell* **157**, 187–200 (2014). doi:10.1016/j.cell.2014.03.003 Medline

4. Ralph Adolphs & David J. Anderson, *The Neuroscience of Emotion A New Synthesis* (Princeton Univ. Press, 2018).

5. R. Adolphs, How should neuroscience study emotions? by distinguishing emotion states, concepts, and experiences. *Soc. Cogn. Affect. Neurosci.* **12**, 24–31 (2017). doi:10.1093/scan/nsw153 Medline

6. J. A. Russell, A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**, 1161–1178 (1980). doi:10.1037/h0077714

7. P. Ekman, An argument for basic emotions. *Cogn. Emotion* **6**, 169–200 (1992). doi:10.1080/02699939208411068

8. J. LeDoux, Rethinking the emotional brain. *Neuron* **73**, 653–676 (2012). doi:10.1016/j.neuron.2012.02.004 Medline

9. L. F. Barrett, A. B. Satpute, Historical pitfalls and new directions in the neuroscience of emotion. *Neurosci. Lett.* **693**, 9–18 (2019). doi:10.1016/j.neulet.2017.07.045 Medline

10. J. Panksepp, The basic emotional circuits of mammalian brains: Do animals have affective lives? *Neurosci. Biobehav. Rev.* **35**, 1791–1804 (2011). doi:10.1016/j.neubiorev.2011.08.003 Medline

11. P. Ekman, Facial expressions of emotion: An old controversy and new findings. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **335**, 63–69 (1992). doi:10.1098/rstb.1992.0008 Medline

12. L. A. Parr, B. M. Waller, J. Fugate, Emotional communication in primates: Implications for neurobiology. *Curr. Opin. Neurobiol.* **15**, 716–720 (2005). doi:10.1016/j.conb.2005.10.017 Medline

13. D. J. Langford, A. L. Bailey, M. L. Chanda, S. E. Clarke, T. E. Drummond, S. Echols, S. Glick, J. Ingrao, T. Klassen-Ross, M. L. Lacroix-Fralish, L. Matsumiya, R. E. Sorge, S. G. Sotocinal, J. M. Tabaka, D. Wong, A. M. J. M. van den Maagdenberg, M. D. Ferrari, K. D. Craig, J. S. Mogil, Coding of facial expressions of pain in the laboratory mouse. *Nat. Methods* **7**, 447–449 (2010). doi:10.1038/nmeth.1455 Medline

14. A. Faure, J. M. Richard, K. C. Berridge, Desire and dread from the nucleus accumbens: Cortical glutamate and subcortical GABA differentially generate motivation and hedonic impact in the rat. *PLOS ONE* **5**, e11223 (2010). doi:10.1371/journal.pone.0011223 Medline

15. K. Finlayson, J. F. Lampe, S. Hintze, H. Würbel, L. Melotti, Facial Indicators of Positive Emotions in Rats. *PLOS ONE* **11**, e0166446 (2016). doi:10.1371/journal.pone.0166446 Medline

16. J. Chikazoe, D. H. Lee, N. Kriegeskorte, A. K. Anderson, Population coding of affect across stimuli, modalities and individuals. *Nat. Neurosci.* **17**, 1114–1122 (2014). doi:10.1038/nn.3749 Medline

17. N. Gogolla, The insular cortex. *Curr. Biol.* **27**, R580–R586 (2017). doi:10.1016/j.cub.2017.05.010 Medline

18. P. Tovote, J. P. Fadok, A. Lüthi, Neuronal circuits for fear and anxiety. *Nat. Rev. Neurosci.* **16**, 317–331 (2015). doi:10.1038/nrn3945 Medline

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, Emotional and social behaviors elicited by electrical stimulation of the insula in the macaque monkey. *Curr. Biol.* **21**, 195–199 (2011). doi:10.1016/j.cub.2010.12.042 Medline

21. C. Fausto, The integration of emotional expression and experience: A pragmatist review of recent evidence from brain stimulation. *Emot. Rev.* **11**, 27–38 (2019). doi:10.1177/1754073917723461

22. D. A. Gehrlach, N. Dolensek, A. S. Klein, R. Roy Chowdhury, A. Matthys, M. Junghänel, T. N. Gaitanos, A. Podgornik, T. D. Black, N. Reddy Vaka, K.-K. Conzelmann, N. Gogolla, Aversive state processing in the posterior insular cortex. *Nat. Neurosci.* **22**, 1424–1437 (2019). doi:10.1038/s41593-019-0469-1 Medline

23. J. Yih, D. E. Beam, K. C. R. Fox, J. Parvizi, Intensity of affective experience is modulated by magnitude of intracranial electrical stimulation in human orbitofrontal, cingulate and insular cortices. *Soc. Cogn. Affect. Neurosci.* **14**, 339–351 (2019). doi:10.1093/scan/nsz015 Medline

24. Y. Peng, S. Gillis-Smith, H. Jin, D. Tränkner, N. J. P. Ryba, C. S. Zuker, Sweet and bitter taste in the brain of awake behaving animals. *Nature* **527**, 512–515 (2015). doi:10.1038/nature15763 Medline

25. L. Wang, S. Gillis-Smith, Y. Peng, J. Zhang, X. Chen, C. D. Salzman, N. J. P. Ryba, C. S. Zuker, The coding of valence and identity in the mammalian taste system. *Nature* **558**, 127–131 (2018). doi:10.1038/s41586-018-0165-4 Medline

26. L. Faget, V. Zell, E. Souter, A. McPherson, R. Ressler, N. Gutierrez-Reed, J. H. Yoo, D. Dulcis, T. S. Hnasko, Opponent control of behavioral reinforcement by inhibitory and excitatory projections from the ventral pallidum. *Nat. Commun.* **9**, 849 (2018). doi:10.1038/s41467-018-03125-y Medline

27. X. Chen, M. Gabitto, Y. Peng, N. J. P. Ryba, C. S. Zuker, A gustotopic map of taste qualities in the mammalian brain. *Science* **333**, 1262–1266 (2011). doi:10.1126/science.1204076 Medline

28. C. Montag, J. Panksepp, Primal emotional-affective expressive foundations of human facial expression. *Motiv. Emot.* **40**, 760–766 (2016). doi:10.1007/s11031-016-9570-x

29. J. M. Susskind, D. H. Lee, A. Cusi, R. Feiman, W. Grabski, A. K. Anderson, Expressing fear enhances sensory acquisition. *Nat. Neurosci.* **11**, 843–850 (2008). doi:10.1038/nn.2138 Medline

30. L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, S. D. Pollak, Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychol. Sci. Public Interest* **20**, 1–68 (2019). doi:10.1177/1529100619832930 Medline

31. N. Dolensek, N. Gogolla, GogollaLab/MouseFacialExpressionAnalysis, Version 1.0, Zenodo (2020). http://doi.org/10.5281/zenodo.3618395

32. J. B. Wekselblatt, E. D. Flister, D. M. Piscopo, C. M. Niell, Large-scale imaging of cortical dynamics during sensory perception and behavior. *J. Neurophysiol.* **115**, 2852–2866 (2016). doi:10.1152/jn.01056.2015 Medline

33. Y. Livneh, R. N. Ramesh, C. R. Burgess, K. M. Levandowski, J. C. Madara, H. Fenselau, G. J. Goldey, V. E. Diaz, N. Jikomes, J. M. Resch, B. B. Lowell, M. L. Andermann, Homeostatic circuits selectively gate food cue responses in insular cortex. *Nature* **546**, 611–616 (2017). doi:10.1038/nature22375 Medline

34. M. P. H. Gardner, A. Fontanini, Encoding and tracking of outcome-specific expectancy in the gustatory cortex of alert rats. *J. Neurosci.* **34**, 13000–13017 (2014). doi:10.1523/JNEUROSCI.1820-14.2014 Medline

35. M. Turk, A. Pentland, Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**, 71–86 (1991). doi:10.1162/jocn.1991.3.1.71 Medline

36. H. S. Dadi, G. K. Mohan Pillutla, Improved face recognition rate using HOG features and SVM classifier. *IOSR J. Electron. Commun. Eng.* **11**, 34–44 (2016). doi:10.9790/2834-1104013444

37. B. A. Sauerbrei, J. Z. Guo, J. D. Cohen, M. Mischiati, W. Guo, M. Kabra, N. Verma, B. Mensh, K. Branson, A. W. Hantman, Cortical pattern generation during dexterous movement is input-driven. *Nature* **577**, 386–391 (2020). 10.1038/s41586-019-1869-9 Medline

38. I. S. Kwak, J.-Z. Guo, A. Hantman, D. Kriegman, K. Branson, Detecting the starting frame of actions in video . arXiv:1906.03340 [cs.CV] (2019).

39. M. Pachitariu *et al.*, Suite2p: beyond 10,000 neurons with standard two-photon microscopy. bioRxiv 061507 [Preprint]. 20 July 2017. https://doi.org/10.1101/061507.

40. A. Miri, K. Daie, R. D. Burdine, E. Aksay, D. W. Tank, Regression-based identification of behavior-encoding neurons during large-scale optical imaging of neural activity at cellular resolution. *J. Neurophysiol.* **105**, 964–980 (2011). doi:10.1152/jn.00702.2010 Medline

41. R. Portugues, C. E. Feierstein, F. Engert, M. B. Orger, Whole-brain activity maps reveal stereotyped, distributed networks for visuomotor behavior. *Neuron* **81**, 1328–1343 (2014). doi:10.1016/j.neuron.2014.01.019 Medline

4. SECOND MANUSCRIPT

From:

Gehrlach, D. A., Dolensek, N., Klein, A. S., Roy Chowdhury, R., Matthys, A., Junghänel, M., Gaitanos, T. N., Podgornik, A., Black, T. D., Reddy Vaka, N., Conzelmann, K. K., & Gogolla, N. (2019). Aversive state processing in the posterior insular cortex. *Nature Neuroscience*, *22*(9), 1424–1437. https://doi.org/10.1038/s41593-019-0469-1

ARTICLES
https://doi.org/10.1038/s41593-019-0469-1

nature neuroscience

Aversive state processing in the posterior insular cortex

Daniel A. Gehrlach[1,2], Nejc Dolensek[1,3], Alexandra S. Klein[1,2], Ritu Roy Chowdhury[1,2], Arthur Matthys[1], Michaela Junghänel[1,5], Thomas N. Gaitanos[1], Alja Podgornik[1,2], Thomas D. Black[1,2], Narasimha Reddy Vaka[1], Karl-Klaus Conzelmann[4] and Nadine Gogolla[1*]

Triggering behavioral adaptation upon the detection of adversity is crucial for survival. The insular cortex has been suggested to process emotions and homeostatic signals, but how the insular cortex detects internal states and mediates behavioral adaptation is poorly understood. By combining data from fiber photometry, optogenetics, awake two-photon calcium imaging and comprehensive whole-brain viral tracings, we here uncover a role for the posterior insula in processing aversive sensory stimuli and emotional and bodily states, as well as in exerting prominent top-down modulation of ongoing behaviors in mice. By employing projection-specific optogenetics, we describe an insula-to-central amygdala pathway to mediate anxiety-related behaviors, while an independent nucleus accumbens-projecting pathway regulates feeding upon changes in bodily state. Together, our data support a model in which the posterior insular cortex can shift behavioral strategies upon the detection of aversive internal states, providing a new entry point to understand how alterations in insula circuitry may contribute to neuropsychiatric conditions.

Ongoing behavior is strongly influenced by affective and homeostatic states, which arise in extended neuronal networks constantly integrating internal and external stimuli[1–3]. Specifically, aversive states signaling threat to the integrity of the body are translated into sustained behavioral adaptation. The posterior, also known as the 'visceral', insular cortex (pIC) is a major cortical convergence site for signals from within and outside the body[3–7]. Human imaging studies and electrophysiological recordings in monkeys and rodents have shown that the pIC receives topographically organized inputs from different internal organs and tissues, transmitting information about changes in heartbeat and the motility of the digestive tract, but also about pain, hunger, nausea, tickle, itch and many more bodily sensations[8–10]. Studies in rodents causally implicated the pIC in gating pain hypersensitivity[11], drug craving and visceral malaise[5]. Furthermore, the pIC integrates information from the external sensory senses[4,12]. Importantly, the responsiveness of the insular cortex (IC) to sensory cues has been shown to depend strongly on the homeostatic state of the animal[13]. While these studies suggest a paramount role for the pIC in processing bodily sensations and homeostatic states, human imaging studies have suggested a second important role for the IC in emotion processing and regulation[4,7,14,15]. Indeed, hyperactivity in the IC is consistently found in individuals with anxiety disorders[16,17]. Furthermore, the IC has been suggested to be a neuroanatomical hub implicated across neuropsychiatric disorders, including depression, addiction, eating disorders and schizophrenia[18–21]. Rodent studies support a role for the pIC in mediating emotional behavior, such as learned fear[22,23], drug–context association[24] and conditioned inhibition of fear[25,26], as well as in mediating approach versus avoidance behavior with respect to social affective stimuli[27].

Taken together, these studies suggest that the pIC lies at the interface between bodily and affective states. However, a comprehensive mechanistic understanding of how this brain region represents internal states or influences motivated behaviors is currently lacking. By using multifaceted anatomical and physiological circuit analyses in mice, we here define a role for the pIC in processing aversive affective and bodily states and regulating behavior via top-down projections to subcortical brain areas.

Results

Optogenetic activation of the pIC drives defensive behavioral and autonomic reactions. To investigate the functional impact of the pIC on behavior, we employed optogenetic tools to broadly manipulate its activity. We first targeted excitatory projection neurons across all layers of the visceral and multisensory areas of the IC[9,13,28] (Fig. 1a and Supplementary Fig. 1). To activate or inhibit insular neurons, we expressed channelrhodopsin-2 (ChR2) or halorhodopsin (eNpHR3.0, abbreviated as NpHR throughout), respectively, fused to enhanced yellow fluorescent protein (eYFP) under the control of the *Camk2a* promoter by using a recombinant adeno-associated virus (AAV) vector and verified that expression of the proteins was specific to glutamatergic projection neurons (Supplementary Fig. 2). Mice injected with the same viral vector carrying only the fluorophore-encoding gene served as controls.

First, we optically stimulated pIC pyramidal neurons in a neutral behavior chamber to assess possible effects of stimulation on locomotion. We chose a 1 s 'on'/4 s 'off' stimulation pattern to avoid overstimulation. This stimulation paradigm did not induce changes in overall locomotion at 10 or 20 Hz (Supplementary Fig. 3a). However, 20-Hz stimulation often led to immediate and highly

[1]Circuits for Emotion Research Group, Max Planck Institute of Neurobiology, Martinsried, Germany. [2]International Max Planck Research School for Molecular Life Sciences, Munich, Germany. [3]Graduate School of Systemic Neurosciences, Ludwig Maximilians University, Munich, Germany. [4]Max von Pettenkofer Institute and Gene Center, Medical Faculty, Ludwig Maximilians University, Munich, Germany. [5]Present address: Ausbildungsinstitut für Kinder- und Jugendlichenpsychotherapie an der Uniklinik Köln (AKiP), Cologne, Germany. *e-mail: ngogolla@neuro.mpg.de

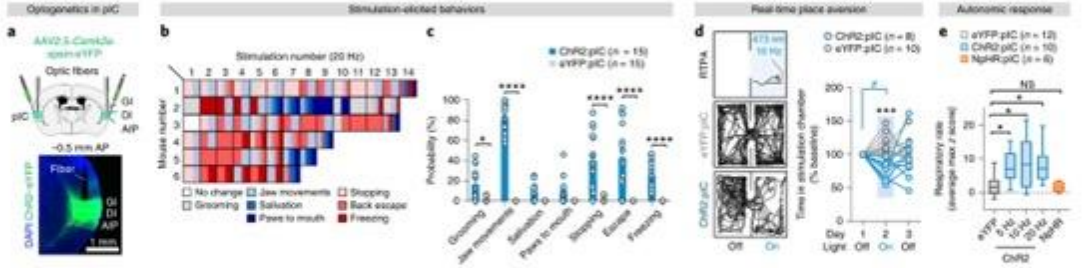


Fig. 1 | pIC activation drives aversive behaviors and avoidance. a, Optogenetic virus injection and optic fiber placement above the pIC. AP, anterior-posterior distance from the bregma; GI, granular insular cortex; DI, dysgranular insular cortex; AIP, posterior agranular insular cortex. **b**, Photostimulation triggers defensive behaviors in ChR2-expressing mice. Rows show reactions of representative individual mice to successive stimulations. Experiments were terminated after 5 min or when severely aversive responses occurred (mice 2–6). **c**, Quantification of light-evoked behavioral responses upon repeated 20-Hz stimulations ($n = 15$ mice per group derived from two independent experiments with similar results; two-way repeated-measures (RM) ANOVA: group (opsin) effect, $F_{(1,28)} = 857.6$, $P < 0.0001$; behavior effect, $F_{(6,168)} = 22.63$, $P < 0.0001$, group × behavior interaction, $F_{(6,168)} = 22.69$, $P < 0.0001$; Bonferroni post hoc analysis: grooming, $^{*}P = 0.0224$; jaw movements, stopping, escape and freezing, $^{****}P < 0.0001$; salivation ($P = 0.8581$) and paws to mouth ($P = 0.5988$) were not significantly altered). **d**, Left: schematic of the RTPA assay and representative locomotor traces of eYFP- and ChR2-expressing mice on the day the laser was on. Right: light activation of pIC neurons results in place aversion in ChR2-expressing mice ($n = 8$ mice expressing ChR2 and $n = 10$ mice expressing only eYFP; two-way RM ANOVA: group (opsin) effect, $F_{(1,16)} = 8.598$, $P = 0.0098$; group × time interaction, $F_{(2,32)} = 4.082$, $P = 0.0264$; Bonferroni post hoc analysis: $^{***}P = 0.0008$ for ChR2-expressing as compared to eYFP-expressing mice with the laser on (day 2); $^{*}P = 0.0463$ for ChR2-expressing mice with the laser off (day 1) versus these mice with the laser on (day 2). **e**, Optogenetic activation increases respiratory rate at all frequencies, while inhibition has no effect (eYFP, $n = 12$ mice (6 mice exposed to blue light and 6 mice exposed to yellow light); ChR2, $n = 10$ mice; NpHR, $n = 6$ mice; one-way ANOVA: $F_{(4,47)} = 4.643$, $P = 0.0033$; Bonferroni post hoc analysis: $^{*}P = 0.0412$ for eYFP-expressing mice versus ChR2-expressing mice stimulated at 5 Hz; $^{*}P = 0.0116$ for eYFP-expressing mice versus ChR2-expressing mice stimulated at 10 Hz; $^{*}P = 0.0234$ for eYFP-expressing mice versus ChR2-expressing mice stimulated at 20 Hz; $P > 0.999$ (NS, not significant) for eYFP-expressing versus NpHR-expressing mice. Bar graphs show mean ± s.e.m. Box-and-whisker plots display the median (center line), 25th to 75th percentile (box), and minimum to maximum values (whiskers).

heterogeneous behavioral reactions that became increasingly frequent upon consecutive stimulations (Fig. 1b,c and Supplementary Video 1). The typical response pattern started with abrupt stopping or freezing; later stimulations also frequently elicited backward escape movements, raising of the front paws to the mouth or salivation. To exclude the possibility that the evoked behaviors might be due to seizure activity, we recorded local field potentials (LFPs) concomitantly with ChR2 stimulation and characterized the LFP activity during all observed behaviors (Supplementary Fig. 4). Although highly frequent and persistent 20-Hz stimulation has the potential to cause seizure-like activity, we here applied much shorter stimulation protocols and verified that all described stimulation-elicited behaviors were not due to seizure-like activity (Fig. 1b,c). Moreover, DREADD-mediated activation of the pIC was found to elicit a subset of the optogenetically elicited reactions, such as freezing and escape movements (Supplementary Fig. 3b). Together, these results demonstrate the potential to evoke aversive and defensive behavioral reactions via pIC activation, while immediate consequences from pIC silencing were not observed (data not shown). Given that stimulation at 20 Hz, but not 10 Hz, elicited immediate reactions, we chose to apply 10-Hz stimulations in all subsequent behavioral tests.

To confirm the apparent adverse effects, we next assessed the valence of the stimulation directly in a real-time place aversion (RTPA) assay. Indeed, upon pIC stimulation, ChR2-expressing mice exhibited strong aversion to the stimulation chamber (Fig. 1d). Interestingly, these mice actively left the stimulation chamber, as shown by an increase in speed in the stimulation chamber, but they did not avoid the stimulation, as they entered the stimulated side as frequently as control mice (Supplementary Fig. 3d). In contrast to stimulation, inhibition had no effect on place preference (Supplementary Fig. 3c).

We noticed that stimulation seemed to increase respiratory rate. Because such an increase could be either a direct consequence of heightened pIC activity or an emotional reaction toward an aversive or arousing stimulation, we repeated pIC photostimulation in anesthetized mice. pIC activation at all tested frequencies increased respiratory rate, whereas NpHR-mediated photoinhibition of the pIC had no effect (Fig. 1e and Supplementary Fig. 3e). We did not observe any change in heart rate upon pIC activation or inhibition (Supplementary Fig. 3f), in contrast to previous electrical stimulation studies in rats and humans[35,36]. Taken together, our results demonstrate that peaks in pIC activity elicit aversive reactions and avoidance, as well as autonomic changes.

Representation of anxiety-related information in the pIC. The above-described findings suggest that increases in pIC activity transmit a negative valence signal, causing animals to exhibit defensive reactions. To address which naturally occurring states or stimuli might evoke such activity and induce spontaneous avoidance, we monitored the activity of pIC neurons by fiber photometry (Fig. 2a) during exploration by mice of an elevated plus maze (EPM) or elevated zero maze (EZM) (Fig. 2e). We first examined how pIC activity varied with occupancy of the open or closed arms of the two mazes. In all mice recorded and in both mazes, we observed much higher levels of pIC activity when the mice occupied the closed arms, while bouts of exploration into the open arms were consistently accompanied by decreased pIC activity (Fig. 2b,c). These data are consistent with high pIC activity during periods of high anxiety, while mice are avoiding exploration and stay in the protected areas of the closed arms. Alternatively, high pIC activity may relate to the safety of the closed compartment. To disentangle these possibilities, we addressed whether the magnitude of closed-arm activity was correlated with the anxiety-related behavior exhibited by individual mice, measured as the time spent in the open arms. Interestingly, we found tight correlations between the anxiety levels of the mice and their pIC activity: in the closed arms, the more anxious the mouse

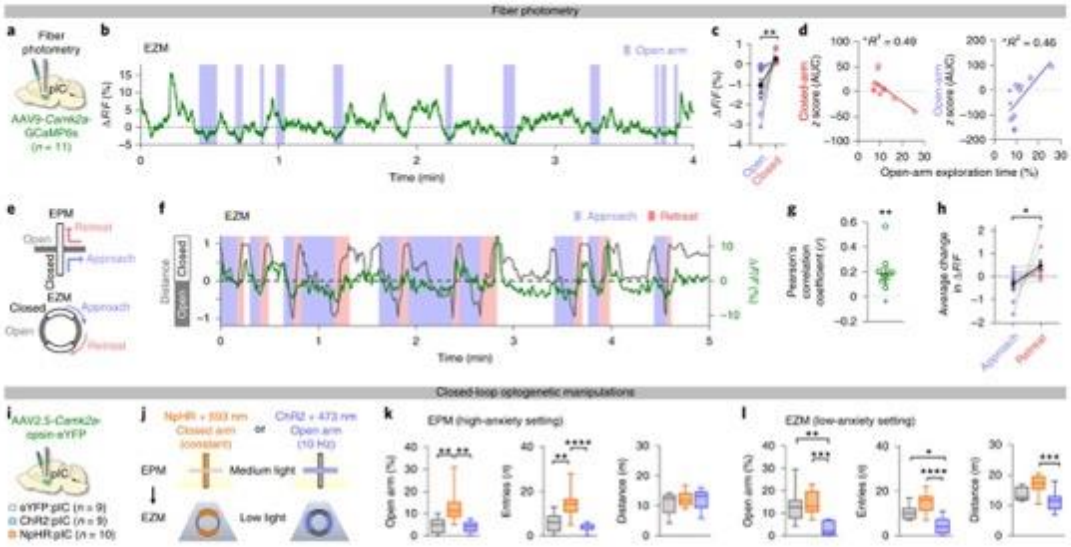


Fig. 2 | The pIC processes and regulates anxiety bidirectionally. a, Optic fiber placement and virus used for fiber photometry. **b**, Example GCaMP6s fluorescence (F) trace from pIC excitatory neurons during 4 min of EZM exploration. Periods in the open arms are shaded in blue. pIC activity decreases in the open arms. **c**, Mean pIC activity is significantly higher in all mice in closed versus open arms in two similar paradigms, the EZM and the EPM ($n = 11$; two-tailed paired t-test: $t = 3.510$, degrees of freedom (d.f.) = 10, $^{**}P = 0.0056$; data combined from 3 mice on the EPM (pluses) and 8 mice on the EZM (circles). The mean ± s.e.m. is shown in black. **d**, Left: mean closed-arm activity as z-scored area under the curve (AUC) in the pIC in both tests correlates with the anxiety of individual mice (time spent in the open arms). Linear regression: $F_{(1,9)} = 8.574$, $^{*}P = 0.0168$, $R^2 = 0.4879$. Right: mean open-arm activity is inversely correlated with anxiety. Linear regression: $F_{(1,9)} = 7.620$, $^{*}P = 0.0221$, $R^2 = 0.4585$. Total $n = 11$ mice (3 in the EPM, shown as pluses, and 8 in the EZM, shown as circles). **e**, Schematic of approach (movement toward the open arms of the mazes) and retreat (movement toward the closed arms of the mazes) behaviors. **f**, Example overlay of pIC activity and distance travelled within the EZM, reflecting approach and retreat behavior. **g**, Correlations between fluorescence and approach or retreat distances are significant for all mice (data not shown), and correlation coefficients (r) are overall significantly different from zero ($n = 11$ mice, including 3 in the EPM (pluses) and 8 in the EZM (circles), as above; two-tailed one-sample t-test: $t = 4.111$, d.f. = 10, $^{**}P = 0.0021$). Error bars show s.e.m. **h**, The average change in fluorescence is negative during approaches and positive during retreats for all mice, resulting in an overall significant difference in the average change in fluorescence between approaches and retreats ($n = 11$ mice; two-tailed paired t-test: $t = 2.255$, d.f. = 10, $^{*}P = 0.0478$). Values from individual mice are connected by gray lines. The mean and s.e.m. are shown in black. **i**, Optogenetic strategy. **j**, Real-time inhibition (left) and stimulation (right) arrangements in the EPM under high-anxiety settings (top) and in the EZM under low-anxiety settings (bottom). **k**, Data derived from $n = 9$ eYFP-, $n = 9$ ChR2- and $n = 10$ NpHR-expressing mice. pIC silencing increases the time spent in the open arms of the EPM (one-way ANOVA: $F_{(2,25)} = 9.267$, $P = 0.0010$; Bonferroni post hoc analysis: $^{**}P = 0.0049$ for NpHR versus eYFP, $^{**}P = 0.0022$ for NpHR versus ChR2, $P > 0.999$ for ChR2 versus eYFP) and the number of entries into the open arms of the EPM (one-way ANOVA: $F_{(2,25)} = 14.99$, $P < 0.0001$; Bonferroni post hoc analysis: $^{**}P = 0.0014$ for NpHR versus eYFP, $^{****}P < 0.0001$ for NpHR versus ChR2, $P = 0.7971$ for ChR2 versus eYFP), but the locomotion distance remains unaltered (one-way ANOVA: $F_{(2,25)} = 0.6660$, $P = 0.5227$). **l**, For the same mice as in **k**, pIC stimulation decreases the time spent in the open arms of the EZM (one-way ANOVA: $F_{(2,25)} = 11.98$, $P = 0.0002$; Bonferroni post hoc analysis: $^{**}P = 0.0019$ for ChR2 versus eYFP, $^{***}P = 0.0004$ for ChR2 versus NpHR, $P > 0.9999$ for NpHR versus eYFP) and the number of entries into the open arms of the EZM (one-way ANOVA: $F_{(2,25)} = 16.29$, $P < 0.0001$; Bonferroni post hoc analysis: $^{*}P = 0.0143$ for ChR2 versus eYFP, $^{****}P < 0.0001$ for ChR2 versus NpHR, $P = 0.0549$ for NpHR versus eYFP) as well as the overall locomotion distance (one-way ANOVA: $F_{(2,25)} = 8.863$, $P = 0.0012$; Bonferroni post hoc analysis: $^{***}P = 0.0009$ for ChR2 versus NpHR, $P = 0.0694$ for NpHR versus eYFP, $P = 0.2954$ for ChR2 versus eYFP). Gray, eYFP-expressing mice; orange, NpHR-expressing mice; blue, ChR2-expressing mice. Box-and-whisker plots display the median (center line), 25th to 75th percentile (box), and minimum to maximum values (whiskers).

was, the higher the closed-arm activity was, while, in the open arms, the more anxious the mouse was, the lower the open-arm activity was (Fig. 2d). When we analyzed the data from the perspective of a mouse's movement in the maze (see Methods; Fig. 2e and Supplementary Fig. 5a,b), we found that pIC activity correlated with the mouse's approach into or retreat from the open arms (Fig. 2f,g). Retreats from the open arms were consistently accompanied by activity increases, whereas approaches were reflected by decreases in activity (Fig. 2h). Notably, we did not find consistent correlations between pIC activity and movement velocity (Supplementary Fig. 5c–e). Together, these data suggest that high pIC activity may prevent exploration of and approach into anxiogenic compartments, while activity decreases permit exploration. Furthermore, these findings raise the possibility that high pIC activity may signal states of anxiety.

pIC bidirectionally controls anxiety-related behaviors. To test directly whether high pIC activity is necessary to maintain anxiety-related behavior, we next examined whether optogenetic activity manipulations in the pIC would cause changes in EPM exploration. Mice were bilaterally injected in the pIC with the same optogenetic viruses described above (Fig. 2i). Informed by the previous activity measurements, we selectively triggered optogenetic silencing of pIC activity when mice occupied the closed arms (Fig. 2j) and activity was expected to be high (Fig. 2c). When compared to controls, mice in which the pIC was silenced entered and spent more time in the

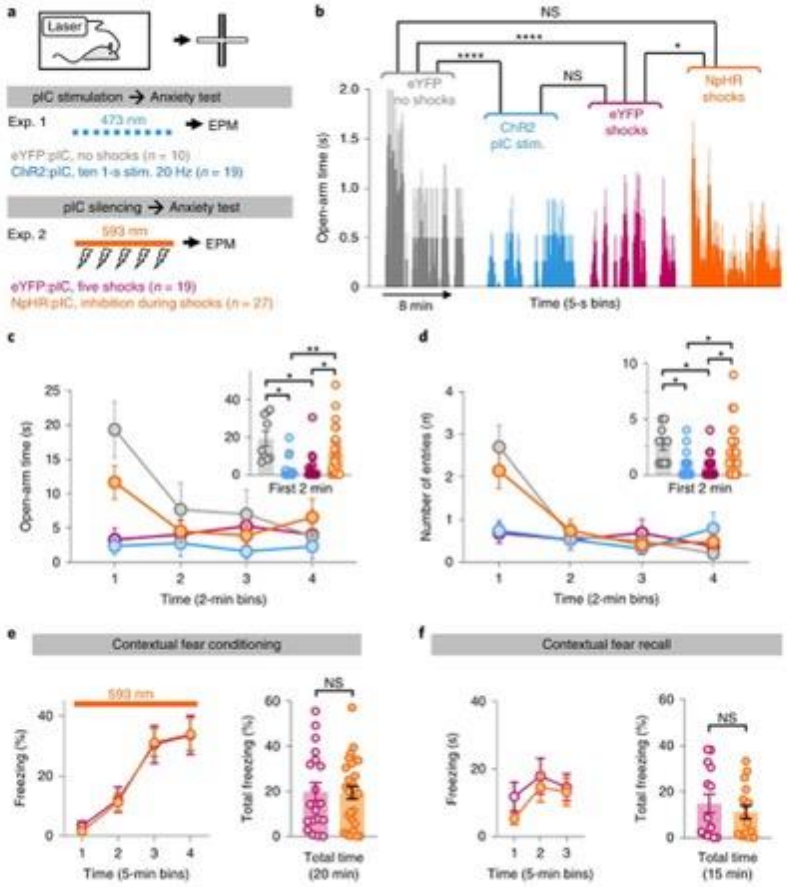


Fig. 3 | The pIC mediates persistent anxiety. a, Illustration of the experimental protocol. Stim., stimulation. **b**–**d**, Data derived from $n = 10$ eYFP-expressing mice without shocks, $n = 19$ ChR2-expressing mice, $n = 19$ eYFP-expressing mice with shocks and $n = 27$ NpHR-expressing mice. **b**, Transient photostimulation of the pIC (blue) decreases the time spent in the open arms during subsequent testing on an EPM as compared to controls (gray). Experience of foot shocks results in a similar reduction in open-arm exploration in eYFP-expressing mice (magenta). This effect is attenuated by pIC inhibition during the shocks (orange). One-way ANOVA: $F_{(3, 380)} = 9.992$, $P < 0.0001$; Bonferroni post hoc analysis: eYFP with shocks versus NpHR, $^{*}P = 0.0481$; eYFP with shocks versus eYFP without shocks, $^{****}P < 0.0001$; eYFP with shocks versus ChR2, $P > 0.9999$; NpHR versus eYFP without shocks, $P = 0.3017$; NpHR versus ChR2, $P = 0.0585$; eYFP without shocks versus ChR2, $^{****}P < 0.0001$. **c**, Analysis of the same data as in **b** in 2-min bins shows that pIC-silenced and no-shock control mice explore the open arms significantly more during the first 2 min than shocked or pIC-stimulated mice. Two-way RM ANOVA: group (opsin) effect, $F_{(3,70)} = 4.340$, $P = 0.0073$; time effect, $F_{(2.678, 187.5)} = 4.502$, $P = 0.0062$; group × time interaction, $F_{(9, 210)} = 2.139$, $P = 0.0277$; Bonferroni post hoc analysis demonstrated significant differences for the following comparisons during the first 2 min: eYFP with shocks versus NpHR, $^{*}P = 0.0415$; eYFP with shocks versus eYFP without shocks, $^{*}P = 0.0244$; NpHR versus ChR2, $^{**}P = 0.0093$; eYFP without shocks versus ChR2, $^{*}P = 0.0173$. Comparisons between eYFP with shocks and ChR2 ($P > 0.9999$) and NpHR versus eYFP without shocks ($P = 0.7770$) were not significant. Inset: individual values for time spent in the open arms during the first 2 min and the significant differences as determined by the Bonferroni test. **d**, Analysis of open-arm entries over time in 2-min bins. Two-way RM ANOVA: group (opsin) effect, $F_{(3,71)} = 1.647$, $P = 0.1863$; time effect, $F_{(2.451, 174)} = 13.9$, $P < 0.0001$; group × time interaction, $F_{(9, 213)} = 3.599$, $P = 0.0003$; Bonferroni post hoc analysis demonstrated significant differences in the first 2 min for the following comparisons: eYFP with shocks versus NpHR, $^{*}P = 0.0262$; eYFP with shocks versus eYFP without shocks, $^{*}P = 0.0221$; NpHR versus ChR2, $^{*}P = 0.0407$; eYFP without shocks versus ChR2, $^{*}P = 0.0273$. Comparisons of eYFP with shocks versus ChR2 and NpHR versus eYFP without shocks were not significantly different (both $P > 0.9999$). Inset: individual values for open-arm entries during the first 2 min and significant differences as determined by the Bonferroni test. **e**, Data from $n = 27$ NpHR- and $n = 19$ eYFP-expressing mice receiving shocks in contextual fear conditioning. Constant pIC inhibition does not affect freezing during the 20-min foot shock period. Left: analysis in 5-min bins shows no difference between groups. Two-way RM ANOVA: group (opsin) effect, $F_{(1,44)} = 0.007057$, $P = 0.9334$. Right: total freezing of all individuals that received foot shocks. Two-tailed unpaired t-test: $t = 0.4193$, d.f. = 13, $P = 0.6818$. **f**, Data from $n = 16$ NpHR- and $n = 14$ eYFP-expressing mice receiving shocks in contextual fear recall. Left: analysis in 5-min bins shows no difference between groups. Two-way RM ANOVA: group (opsin) effect, $F_{(1,24)} = 0.6022$, $P = 0.4442$. Right: total freezing of all individuals that received foot shocks. Two-tailed unpaired t-test: $t = 0.7586$, d.f. = 28, $P = 0.4544$. Bar and line graphs show mean ± s.e.m. Box-and-whisker plots display the median (center line), 25th to 75th percentile (box), and minimum to maximum values (whiskers).

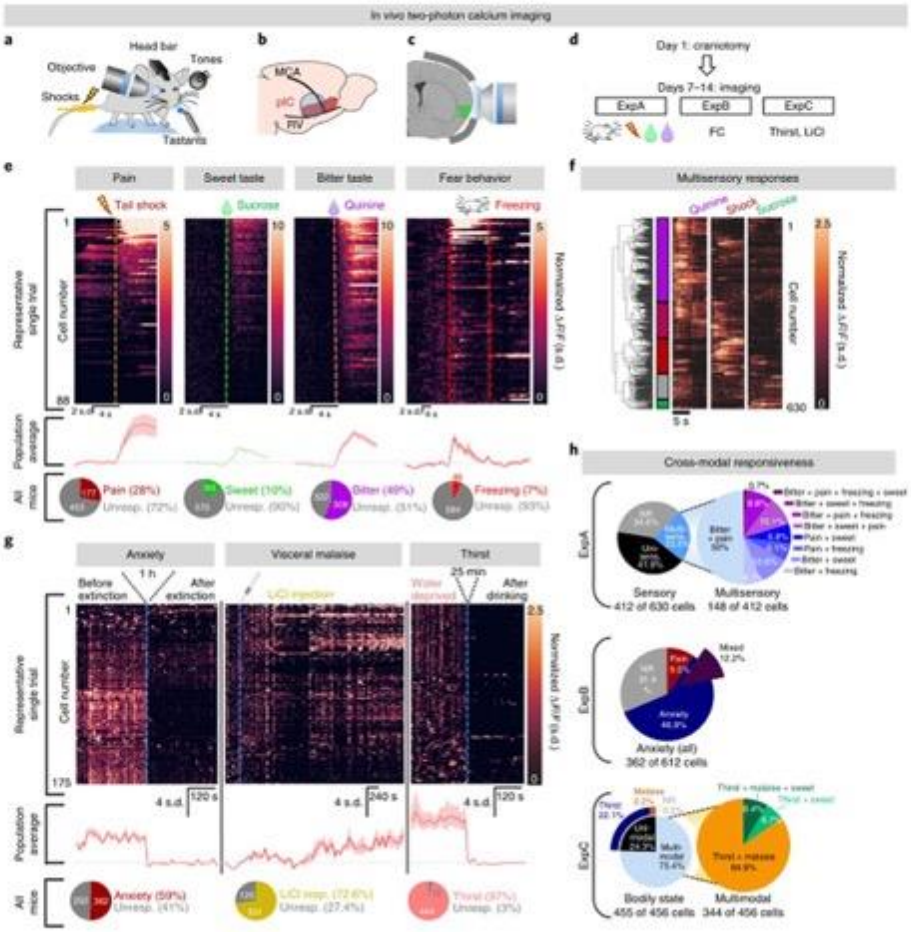


Fig. 4 | Sensory, emotional and bodily stimuli elicit insular activity. a, Two-photon calcium imaging of the pIC in awake, head-fixed mice. **b**, Schematic showing positioning of the chronic window implant (blue circle) above the pIC (red) on the lateral brain surface with respect to two major blood vessels, the medial cerebral artery (MCA) and the rhinal vein (RV). **c**, Imaging window position illustrated on a coronal section. **d**, The experimental timelines for three sets of experiments (ExpA–ExpC). FC, fear conditioning. **e**, pIC responsiveness to acute stimuli. Top: heat maps (not binned, 7.7 frames per second (f.p.s.)) showing response amplitudes (normalized $\Delta F/F$ = s.d.) of 88 individual neurons from a single region of interest (ROI) in a representative mouse. Rows are cells sorted by responsiveness to a tail shock (left), sweet (sucrose) or bitter (quinine) tastants (middle) or a freezing episode (right) and aligned to stimulus onset (dashed line). Middle: average $\Delta F/F$ (s.d.) traces of all neurons with significantly higher activity during the aversive state. Bottom: pie charts showing proportions of stimulus-responsive neurons across all mice (cell counts per population are reported in the charts). $n = 630$ neurons from $n = 3$ mice. Throughout, all cells that responded to quinine, sucrose and water (6.8%, 48 of 630 cells) were assumed to encode for orofacial sensory experience of drinking and were excluded from the quinine- and sucrose-responding cell populations. Unresp., unresponsive. **f**, Multisensory responses of all recorded neurons. Heat maps (7.7 f.p.s.) show the response profiles of 630 individual neurons from $n = 3$ mice to quinine, tail shock and sucrose, directly after the start of stimulus presentation. Neurons were sorted on the basis of their response profile across all three stimuli by hierarchical clustering (dendrogram displaying clusters shown to the left). A graphic representation of the identified functional clusters of neurons based on response profiles is shown to the right of the dendrogram (purple, bitterness-responding neurons; red, pain-responding neurons; purple and red, pain- and bitterness-responding neurons; gray, non-responsive neurons; green, sweetness-responding neurons). Note the large number of neurons responding across two negatively valent stimuli. **g**, pIC responses to affective and bodily states. Top: heat maps showing response amplitudes of 175 individual neurons from a single ROI in one example mouse per condition (1 pixel = ~13 s = 100 frames binned for anxiety and thirst; ~32.5 s = 250 frames binned for malaise). Rows correspond to cells clustered by response profile toward states of high versus low anxiety (left), neutral versus sick body feeling (middle) and thirsty versus water-satiated state (right). Middle: average $\Delta F/F$ (s.d.) traces of all neurons with significantly higher activity during the aversive state. Bottom: pie charts showing proportions of stimulus-responsive neurons across all mice ($n = 3$ mice per condition; cell counts per population are reported in the charts). **h**, Quantification of multisensory and multimodal overlaps in all cells recorded within each of the three experiments. Top: ExpA (high proportion of multisensory cells); middle: ExpB (more than half of all neurons were modulated by anxiety); bottom: ExpC (>99% of all recorded cells were modulated by bodily state (thirst and/or malaise)). Three-quarters of all cells were mixed coding. NR, non-responsive. Line graphs show mean ± s.e.m. (shaded).

open arms of the EPM without changing their locomotion, in line with an anxiolytic effect (Fig. 2k). Similar effects were also observed when the pIC was silenced via an inactivating DREADD strategy in an open-field test (Supplementary Fig. 6d,e). However, optogenetic pIC inhibition in alternating 2-min laser on/off epochs was not sufficient to robustly induce anxiolysis, and only a trend toward increased time spent in the open arms was observed (Supplementary Fig. 6b). Together, these data suggest that pIC silencing is efficient when sustained for long periods of time.

We next assessed whether selectively optogenetically increasing pIC activity while mice explored the open arms, where pIC activity was expected to be low, would result in retreat from these compartments. We did not observe any difference in exploration between eYFP- and ChR2-expressing mice in the EPM (Fig. 2k). This lack of effect may be due to the already high anxiety levels under the chosen settings. We therefore tested the same mice under low-anxiety settings in an equivalent test, the EZM (Fig. 2j). Indeed, in the EZM, ChR2-mediated pIC activation in the open arms resulted in stronger avoidance, as demonstrated by fewer entries into and less time spent in the open arms (Fig. 2l), corresponding to an anxiogenic effect. Interestingly, under these low-anxiety settings, pIC silencing was not able to significantly increase open-arm exploration further. Because we observed real-time place aversion upon pIC stimulation (Fig. 1d), the observed effects in the EZM could be due to aversion rather than increased anxiety. To address this possible confounder, we performed optogenetic pIC stimulation in a 2-min laser on/off paradigm and in a separate experiment with an activating DREADD strategy in the EPM. Both experiments confirmed that enhancing pIC activity resulted in increased anxiety-like behavior (Supplementary Fig. 6c,f).

Taken together, our results demonstrate that the pIC exerts a strong and bidirectional influence on anxiety-related behaviors. Furthermore, owing to the slow emergence and sustained effects observed, our data collectively hint at a role for the pIC in processing anxiety states.

Activity in the pIC is sufficient and necessary to mediate sustained anxiety. The experiments described above showed dynamic transitions in pIC activity between approach and avoidance behavior and thus possibly in the spontaneous transitions between low- and high-anxiety states. To explore more directly whether the pIC is involved in the emergence of anxiety states, we tested whether transient activity manipulations in the pIC had sustained behavioral effects.

First, we applied ten 1-s-long photostimulations at 20 Hz to the pIC of ChR2- or eYFP-expressing mice in an open field. Immediately following this, we tested the animals' behavior on the EPM without further activity manipulations (Fig. 3a). Intriguingly, following this treatment, ChR2-expressing mice entered the open arms less frequently and spent less time there than eYFP-expressing controls (Fig. 3b,c), especially during the first 2 min of the experiment (Fig. 3c,d). Thus, transient pIC stimulation induced a sustained anxiety-like state.

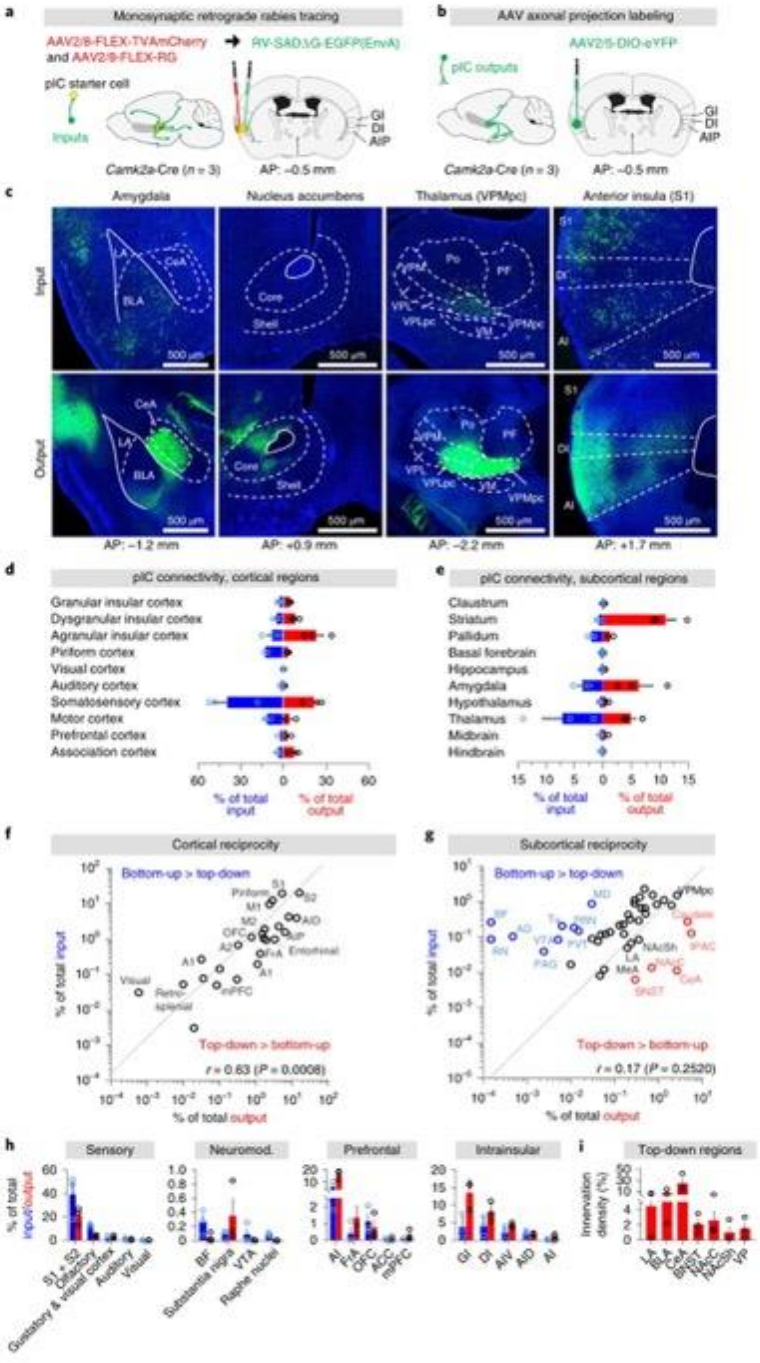
We next addressed the necessity of the pIC in the emergence of anxiety states. To this end, mice underwent unsignaled foot shocks in a conditioning chamber. In one group of mice, the pIC was silenced through constant NpHR illumination during the foot shock period. The efficiency of long-term inhibition was confirmed by optrode recordings (Supplementary Fig. 7a–c). When subsequently tested on the EPM (Fig. 3b–d), eYFP-expressing control mice having experienced the foot shocks exhibited a significant decrease in open-arm exploration when compared to non-shocked controls (Fig. 3c,d). NpHR-expressing mice exhibited a significant rescue of this phenotype (Fig. 3b–d), behaving similarly to non-shocked eYFP-expressing controls. Conversely, previously ChR2-stimulated mice behaved very similarly to controls that had received foot shocks.

To exclude the possibility that the effects of pIC silencing were due to a reduction in pain or fear experienced during the foot shock period, we compared freezing, as a behavioral measure of fear, expressed during the shock period and several days later in the conditioning chamber during contextual fear recall. We found that freezing was not different either during foot shocks (when pIC was constantly silenced) or during recall of the contextual conditioning between eYFP- and NpHR-expressing mice (Fig. 3e,f). These results were confirmed in an independent auditory fear conditioning experiment performed under constant pIC inhibition, which showed no effects during the conditioning or subsequent fear recall from the context or the conditioned stimuli (Supplementary Fig. 8b–d). Furthermore, we verified that no differences occurred in the pain thresholds of pIC-inhibited mice in a hot plate test (Supplementary Fig. 8a). In sum, these data suggest that pIC inhibition does not impair acute reactions to painful stimuli or the expression of fear through freezing or prevent context or cue associations. Instead, they support the notion that the pIC has an important role in the emergence of sustained states of anxiety.

Representation of aversive stimuli and states in single neurons of the pIC. Our data collectively suggest a role for the pIC in mediating aversive emotion states such as anxiety. However, it is unclear how negative emotion states arise and are processed within the IC. Bodily as well as sensory stimuli are potent triggers of affective states, and the IC is known to receive highly diverse information streams[4,9,29]. We therefore sought to investigate how diverse stimuli

Fig. 5 | Whole-brain tracings of direct pIC inputs and outputs. a,**b**, Schematics of the monosynaptic retrograde (**a**) and anterograde (**b**) tracing strategies employed to identify pIC inputs and outputs. **c**, Example images of input and output patterns in representative pIC-connected brain regions. Findings were replicated in three other mice for each input and output. **d**–**i**, Data from $n = 3$ mice for inputs and $n = 3$ mice for outputs. **d**,**e**, Direct comparison between inputs and outputs of pIC projection neurons from larger brain areas, divided into cortical (**d**) and subcortical (**e**) connectivity. **f**,**g**, Analysis of the reciprocity of input and output connectivity between single brain regions and the pIC (25 data points representing cortical brain regions (**f**) and 49 data points representing subcortical brain regions (**g**) are mean values from $n = 3$ mice). **f**, Most cortical regions have reciprocal connections with the pIC, as shown by the significant correlation between inputs and outputs ($n = 25$, Pearson's $r = 0.63$ (95% confidence interval (CI) = 0.31 to 0.82); linear regression showed significant deviation from zero, two-tailed $P = 0.0008$). **g**, Many subcortical regions do not show strong reciprocity. For some regions, the pIC inputs are dominating (top-down > bottom-up; red), while for others the inverse is true (bottom-up > top-down; blue). Altogether, the input and output were not correlated for subcortical regions ($n = 49$, Pearson's $r = 0.17$ (95% CI = −0.12 to 0.43); linear regression showed no significant deviation from zero, two-tailed $P = 0.2520$). **h**, Analysis of inputs and outputs from functionally grouped regions, such as sensory, neuromodulatory, prefrontal and intra-insular areas. **i**, Density of pIC innervation in select top-down target regions. ACC, anterior cingulate cortex; AD, anterodorsal thalamic nucleus; AIC, agranular insular cortex; AID, dorsal agranular insular cortex; AIP, posterior agranular insular cortex; AIV, ventral agranular insular cortex; BNST, bed nucleus of the stria terminalis; CeA, central amygdala; DI, dysgranular insular cortex; GI, granular insular cortex; FrA, frontal association cortex; IPAC, interstitial nucleus of the posterior limb of the anterior commissure; MD, mediodorsal thalamus; mPFC, medial prefrontal cortex; NAcC, nucleus accumbens core; OFC, orbitofrontal cortex; PBN, parabrachial nucleus; PVT, paraventricular thalamic nucleus; RN, raphe nuclei; Tu, olfactory tubercle; VTA, ventral tegmental area; VP, ventral pallidum. Bar graphs show mean ± s.e.m.

relevant to affective and bodily states might be processed in single neurons of the pIC. To this end, we developed an approach for cellular-resolution imaging of the pIC in awake mice using two-photon microscopy (Fig. 4a). The field of view was positioned above the posterior half of the IC, including visceral, multisensory and gustatory regions (Fig. 4b). Expression of the calcium indicator GCaMP6s



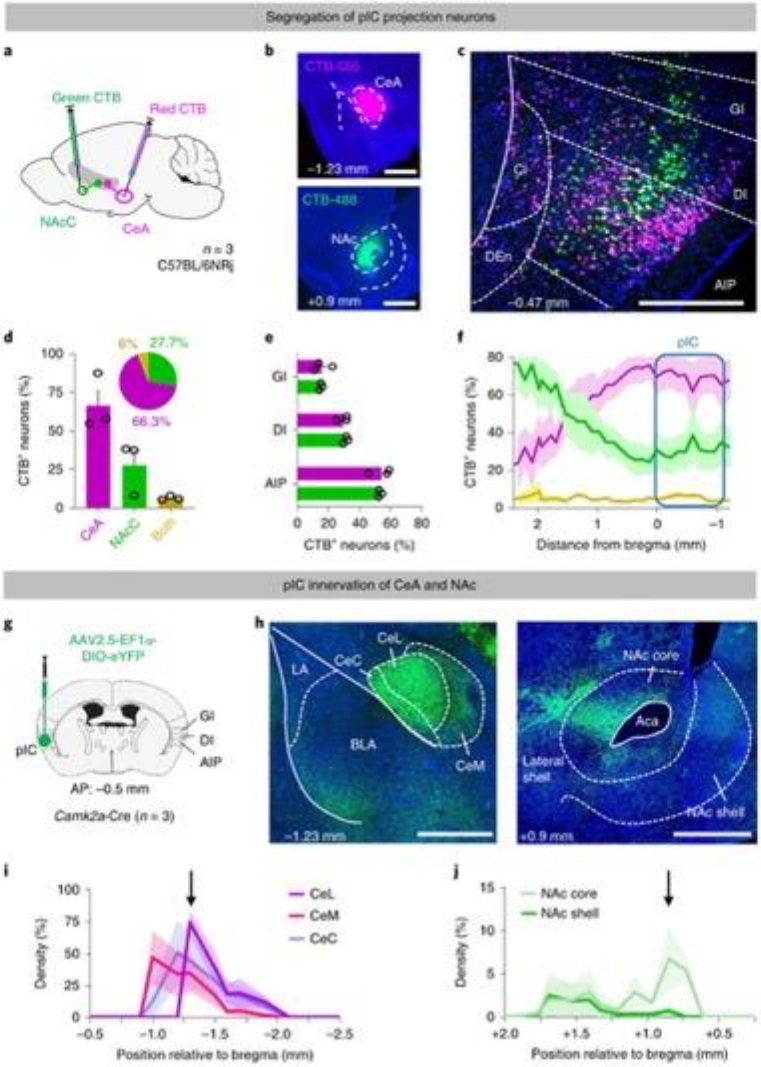


Fig. 6 | Largely non-overlapping pIC neuronal subpopulations project to the CeA and NAcC. a, CTB retrograde labeling of pIC projections to the CeA (CTB-555) and NAcC (CTB-488). **b**, Representative images of CTB injection sites in the CeA (top) and NAcC (bottom). Similar results were obtained in three mice. **c**, Representative image of retrograde labeling of pIC neurons projecting to the CeA (magenta), NAcC (green) or both (yellow). **d**–**f**, Data from $n = 3$ mice. **d**, Quantification of retrograde labeling of pIC neurons. pIC→CeA projections (4,863 ± 996.7 (s.e.m.) cells in total) and pIC→NAcC projectors (2,050 ± 642.0 cells in total) were mostly non-overlapping, with <6% of cells (405.7 ± 21.46 cells) projecting to both. **e**, The two populations, when normalized to total cell counts, were distributed similarly in the different pIC subregions (absolute cell counts (mean ± s.e.m.): GI: CeA projections, 702.7 ± 508.2; NAc projections,120.7 ± 25.9; DI: CeA projections, 1,161.0 ± 572.4; NAc projections, 310 ± 70.3; AIP: CeA projections, 2,051.3 ± 741.1; NAc projections, 618 ± 102.8). **f**, Distribution of CeA and NAcC projectors along the anterior–posterior axis of the insular cortex (+2.4 to −1.22 mm with respect to the bregma). In the anterior insula, we found more NAcC projectors than CeA projectors. In contrast, in the posterior insula (blue box), CeA projectors were more abundant than NAcC projectors. **g**, Schematic of anterograde labeling of pIC axons. **h**, Representative images of eYFP-labeled pIC axons in the amygdala (left) and nucleus accumbens (right). Note the strong innervation within patches in the NAcC. Similar results were obtained in three mice. **i**,**j**, Quantification of the innervation densities of pIC axons ($n = 3$ mice) in the amygdala (**i**) and nucleus accumbens (**j**) along the anterior–posterior axis, identifying the sites with the highest innervation densities. Sites of optic fiber implant for the experiments described in Figs. 7 and 8 are marked with arrows. Line and bar graphs show mean ± s.e.m. Scale bars, 500 µm. GI, granular insular cortex; DI, dysgranular insular cortex; AIP, posterior agranular insular cortex; CeC, central amygdala, capsular part; CeL, central amygdala, lateral part; CeM, central amygdala, medial part; DEn, dorsal endopirifrom nucleus; Cl, claustrum; LA, lateral amygdala; BLA, basolateral amygdala.

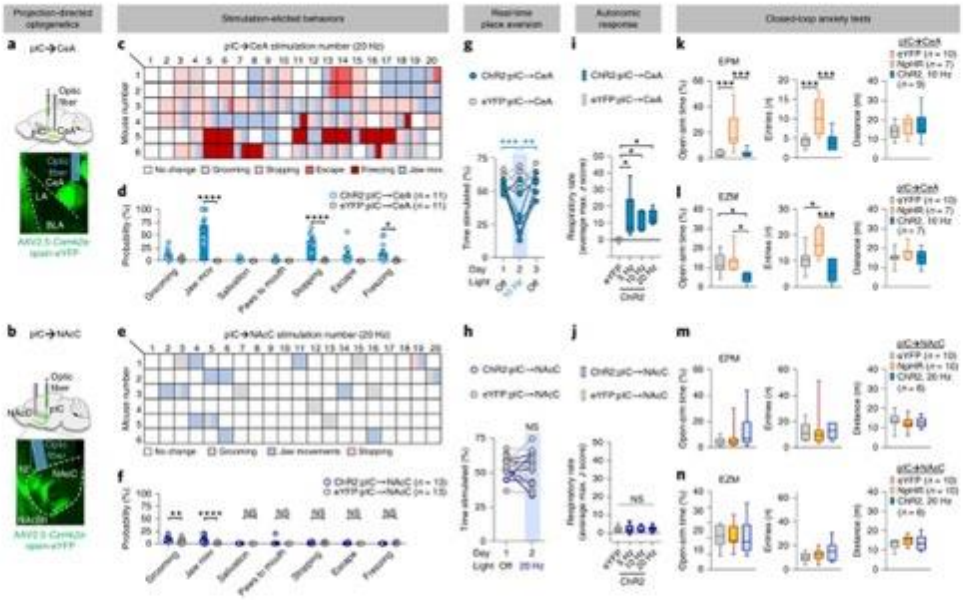


Fig. 7 | pIC input to the CeA governs defensive reactions, avoidance and anxiety-related behaviors. a,**b**, Viral injection and optic fiber placement for pIC→CeA (**a**) and pIC→NAcC (**b**) projection-specific optogenetics. **c**–**f**, Stimulation-elicited behavioral reactions upon 20-Hz stimulation of the pIC→CeA or pIC→NAcC pathway. **c**,**e**, Reactions over time to successive stimulations in six representative mice for each pathway. Stimulations of the pIC→CeA pathway elicited prominent reactions (**c**), whereas very few stimulations of the pIC→NAcC pathway elicited reactions (**e**). **d**, Quantification of the reactions upon pIC→CeA stimulation in $n = 11$ mice per group. Two-way RM ANOVA: group (opsin) effect, $F_{(1,20)} = 92.41$, $P < 0.0001$; behavior effect, $F_{(6, 120)} = 28.64$, $P < 0.0001$; group × behavior interaction, $F_{(6, 120)} = 29.30$, $P < 0.0001$; Bonferroni post hoc analysis demonstrated significant differences between eYFP- and ChR2-expressing mice for jaw movements and stopping (both ****$P < 0.0001$) and freezing (*$P = 0.0126$). No differences were found for grooming ($P = 0.3727$), paws to mouth and salivation (both $P > 0.9999$), while a trend was observed for escapes ($P = 0.0503$). **f**, Quantification of the reactions upon pIC→NAcC stimulation in $n = 13$ mice per group. Two-way RM ANOVA: group (opsin) effect, $F_{(1,24)} = 12.67$, $P = 0.0016$; behavior effect, $F_{(6, 144)} = 24.82$, $P < 0.0001$; group × behavior interaction, $F_{(6, 144)} = 5.282$, $P < 0.0001$; Bonferroni post hoc analysis demonstrated significant differences between eYFP- and ChR2-expressing mice for grooming (**$P = 0.0095$) and jaw movements (****$P < 0.0001$). No differences were observed for escapes, paws to mouth, salivation, freezing or stopping (all $P > 0.9999$). **g**, pIC→CeA stimulation results in real-time place aversion behavior ($n = 7$ ChR2- and $n = 6$ eYFP-expressing mice; two-way RM ANOVA demonstrated significant differences for group (opsin) effect ($F_{(1,11)} = 21.15$, $P = 0.0008$) and laser effect ($F_{(2,22)} = 5.531$, $P = 0.0113$) as well as for opsin × laser interaction ($F_{(2,22)} = 5.368$, $P = 0.0126$); Bonferroni post hoc analysis demonstrated significant differences between habituation (day 1, laser off) and the laser-on day (day 2) (***$P = 0.008$) as well as between the laser-on day (day 2) and the memory test (day 3, laser off) (**$P = 0.0017$) for the ChR2-expressing group but no difference for the eYFP-expressing group (comparisons between light-on and light-off days, $P > 0.9999$)). eYFP- and ChR2-expressing mice exhibited a different preference only under light stimulation ($P < 0.0001$). **h**, Laser stimulation of the pIC→NAcC pathway does not influence place preference ($n = 7$ ChR2- and $n = 10$ eYFP-expressing mice; two-way RM ANOVA demonstrated no significant differences for group (opsin) effect ($F_{(1,15)} = 0.5608$, $P = 0.4655$), light effect ($F_{(1,30)} = 1.208$, $P = 0.2890$) or opsin × laser interaction ($F_{(1,15)} = 0.6262$, $P = 0.441$)). **i**, pIC→CeA stimulation induces increases in respiratory rate in anesthetized mice ($n = 5$ ChR2- and $n = 8$ eYFP-expressing mice; one-way ANOVA: $F_{(3,99)} = 5.098$, $P = 0.0093$; Bonferroni multiple-comparisons test for comparison to eYFP-expressing controls: 5 Hz, *$P = 0.0214$; 10 Hz, *$P = 0.0296$, 20 Hz, *$P = 0.0172$). **j**, pIC→NAcC stimulation does not affect respiratory rate (ChR2: $n = 6$ mice stimulated at 5 and 20 Hz and $n = 4$ mice stimulated at 10 Hz; eYFP: $n = 9$ mice; one-way ANOVA: $F_{(3,25)} = 0.02645$, $P = 0.9940$). **k**–**n**, Assessment of effect of pathway-specific optogenetics on anxiety. **k**, pIC→CeA inhibition has anxiolytic effects in the EPM ($n = 10$ eYFP-, $n = 7$ NpHR- and $n = 9$ ChR2-expressing mice). Open-arm time: one-way ANOVA: $F_{(2,23)} = 11.87$, $P = 0.0003$; Bonferroni post hoc analysis: NpHR versus eYFP, ***$P = 0.0006$; ChR2 versus eYFP, $P > 0.9999$; NpHR versus ChR2, ***$P = 0.0009$. Open-arm entries: one-way ANOVA: $F_{(2,23)} = 13.48$, $P = 0.0001$; Bonferroni post hoc analysis: NpHR versus eYFP, ***$P = 0.0004$; ChR2 versus eYFP, $P = 0.9704$; NpHR versus ChR2, ***$P = 0.0003$. Distance travelled in the entire maze was not affected by optogenetic manipulations of pIC→CeA: one-way ANOVA: $F_{(2,23)} = 0.6430$, $P = 0.5349$. **l**, Stimulation of pIC→CeA has anxiogenic effects in the EZM, while inhibition only increases entries into but not time spent in the open arms ($n = 10$ eYFP-, $n = 7$ NpHR- and $n = 7$ ChR2-expressing mice). Open-arm time: one-way ANOVA: $F_{(2,21)} = 6.343$, $P = 0.0070$; Bonferroni post hoc analysis: NpHR versus eYFP, $P > 0.9999$; ChR2 versus eYFP, *$P = 0.0221$; NpHR versus ChR2, *$P = 0.0111$. Open-arm entries: one-way ANOVA: $F_{(2,21)} = 9.533$, $P = 0.0011$; Bonferroni post hoc analysis: NpHR versus eYFP, *$P = 0.0276$; ChR2 versus eYFP, $P = 0.2498$; NpHR versus ChR2, ***$P = 0.0009$. Distance travelled in the entire maze is not affected by optogenetic manipulations of pIC→CeA: one-way ANOVA: $F_{(2,21)} = 0.9968$, $P = 0.3859$. **m**,**n**, Optogenetic pIC→NAcC manipulations do not affect anxiety-related behaviors in the EPM or EZM ($n = 10$ eYFP-, $n = 10$ NpHR- and $n = 6$ ChR2-expressing mice). **m**, EPM open-arm time: one-way ANOVA: $F_{(2,23)} = 1.465$, $P = 0.2519$. EPM open-arm entries: one-way ANOVA: $F_{(2,23)} = 0.01173$, $P = 0.9883$. Distance travelled in the EPM: one-way ANOVA: $F_{(2,23)} = 0.5072$, $P = 0.6088$. **n**, EZM open-arm time: one-way ANOVA: $F_{(2,23)} = 0.006972$, $P = 0.9931$. EZM open-arm entries: one-way ANOVA: $F_{(2,23)} = 1.544$, $P = 0.2349$. Distance travelled in the EZM: one-way ANOVA: $F_{(2,23)} = 1.244$, $P = 0.3068$. Bar graphs show mean ± s.e.m. Box-and-whisker plots display the median (center line), 25th to 75th percentile (box), and minimum to maximum values (whiskers).

(see Methods) allowed us to simultaneously and repeatedly image the activity of 100–200 neurons in layer 2/3 of each mouse (Fig. 4c). The imaging was divided into three different experiments (Fig. 4d; see Methods).

We first examined neural responses to sensory stimuli of strong innate valence, such as painful stimuli (tail shocks) and pleasant or unpleasant tastants (sucrose and quinine, respectively). We found substantial fractions of pIC neurons responding readily to tail shocks or quinine (28% and 49%, respectively), whereas cells responding to sucrose were more sparse (10%) (Fig. 4e), in line with previous studies[28,31].

To address whether responsiveness of insular neurons was also elicited by the behavioral expression of emotions, we determined the activity of insular neurons during freezing responses, which occurred spontaneously following the delivery of tail shocks (see Methods for details on freezing detection and Supplementary Fig. 9a–c). We found that a small fraction (7%) of pIC neurons was activated by freezing (Fig. 4e).

Interestingly, when analyzed across different sensory stimuli, many pIC neurons were found to be multisensory. In particular, cells responding to both quinine and tail shocks were observed (Fig. 4f).

Given our results implicating the pIC in regulating anxiety, we investigated how anxiety is represented within the pIC on a single-cell level. To this end and to bring mice into high- or low-anxiety states, we compared the spontaneous activity of insular neurons 1 h after fear conditioning (high anxiety, referred to as 'before extinction') or directly after fear extinction learning (low anxiety). Because mice had their heads fixed, the context stayed the same during the entire fear and extinction procedure. We therefore used tones as conditioned stimuli and extinction cues to achieve better extinction results. We found a marked increase in the spontaneous activity of insular neurons 1 h after fear conditioning as compared to habituation or extinction (Fig. 4g, left, and Supplementary Fig. 10a). These results were not due to fluorescence bleaching or habituation of the mice, as we imaged throughout the sessions and pIC activity increased with fear conditioning but decreased with extinction learning (Supplementary Fig. 10a,g). Furthermore, although fluorescence was also decreased upon contextual extinction (no tones), activity decreases were not as pronounced as when conditioned auditory stimuli were used to extinguish fear, in agreement with the idea that global pIC activity levels scaled with anxiety levels (Supplementary Fig. 10b).

Given the known function of the pIC as a viscerosensitive cortex, we next examined how changes in bodily states are represented in the pIC. We used lithium chloride (LiCl) injections to induce malaise[31] or water deprivation to elicit thirst, and compared the spontaneous activity of insular neurons before and after these manipulations. We found increased activity in large fractions of insular neurons during these aversive bodily states (malaise and thirst) when compared to equivalent baseline periods (before LiCl injection or after drinking) (Fig. 4g, middle and right), extending earlier c-Fos studies[4,32]. Notably, we observed activity changes in opposite directions when aversive states emerged (LiCl) and subsided (after drinking or extinction; Fig. 4g) and no increased activity upon intraperitoneal injections of saline solution (Supplementary Fig. 10c), arguing for highly specific correlations between high pIC activity and internal states.

When using pupillometry as a measure of general arousal[33], we detected no strong correlations between pupil diameter and neuronal firing in the pIC (Supplementary Fig. 9d), arguing that increases in pIC activity during aversive states cannot be explained by arousal.

To assess the extent to which single neurons in the pIC responded to more than one sensory stimulus or internal state, we quantified the overlap of responses in the three different experiments. We found that 23.5% of single neurons in the pIC were multisensory, responding to more than one of the acute stimuli tested (Fig. 4h, top); >50% of cells were modulated by anxiety, some of which (12.2%) were also responsive to painful tail shocks (Fig. 4h, middle). Notably, >99% of cells were modulated by the bodily state of the mice, such as either thirst or malaise, and 75% of these cells were mixed coding, exhibiting responsiveness to both states or to a state and sucrose (Fig. 4h, bottom). Together, these data suggest a very high level of convergence and integration of acute sensory stimuli and sustained affective and homeostatic states within the pIC.

Comprehensive pIC input and output mapping identifies inputs from diverse functional systems as well as prominent top-down projections. The observed roles of the pIC in processing aversive states and in regulating motivated behaviors prompted us to explore the neuronal circuitry that, first, could give rise to such responsiveness (pIC inputs) and, second, could mediate the observed behavioral modulation (pIC outputs).

To explore direct monosynaptic inputs to the pIC, we employed cell-type-specific retrograde rabies tracings[34] of CamKIIα-positive neurons in the pIC (Fig. 5a and Supplementary Fig. 11). To explore the axonal target regions of pIC neurons, we injected AAVs with Cre-dependent expression of eYFP into *Camk2a*-Cre mice (Fig. 5b). Our results confirmed and extended earlier studies using conventional tracers in different species showing that the pIC receives highly multimodal and associative inputs[8,9,35,36]. Comprehensive whole-brain analysis (Fig. 5c–i and Supplementary Fig. 12) showed that the pIC integrates inputs from a large set of areas, including sensory, autonomic, motor, associative and limbic structures, which may underlie its responsiveness to highly distinct stimuli and states. Furthermore, it sends outputs to a similarly large and heterogeneous collection of target regions (Fig. 5d,e and Supplementary Fig. 12). When focusing on larger categories of information streams, we found that the pIC was marked by intensive intra-insular connectivity (Fig. 5d,h). Furthermore, the pIC received heavy sensory input from both cortical and thalamic sources (Fig. 5e–h) and sent its strongest subcortical output to the striatum, the amygdala and the thalamus (Fig. 5e,g,i). Next, we analyzed the reciprocity of inputs and outputs. Interestingly, cortical connectivity was highly reciprocal, as demonstrated by strong correlation of inputs and outputs (Fig. 5f), suggesting prominent associative properties. In contrast, subcortical connectivity was for many areas not reciprocated and overall was not correlated (Fig. 5g). Indeed, some regions were prominently biased toward being mostly top-down projection targets, such as the central amygdala (CeA), the core of the nucleus accumbens (NAcC) and the bed nucleus of the stria terminalis (BNST). Others mostly sent bottom-up input, such as the raphe nuclei, the ventral tegmental area (VTA) and several thalamic nuclei (Fig. 5g). Finally, we clustered the input and output connectivity regions into functional groups. This analysis highlighted the multisensory inputs and diverse neuromodulatory drive, as well as the strong connectivity within the insula and with other regions of the prefrontal cortex (Fig. 5h). Taken together, our comprehensive input tracings identify various information streams from sensory, limbic and associative brain regions that may underlie the highly complex responses observed in the pIC. When comparing the projection densities of axons in potential top-down target regions (Fig. 5i), we identified the CeA and NAcC among the subcortical target regions receiving dense innervation (confined to reproducible patches in the NAcC; Figs. 5c and 6h). Because both the amygdaloid complex and nucleus accumbens are major brain hubs regulating positively and negatively motivated behaviors and have been implicated in anxiety, aversion and reward[37–38], we hypothesized that the pIC→CeA and pIC→NAcC pathways might represent important candidates to mediate pIC top-down control over motivated behaviors.

Separate pIC neurons project to the CeA and NAcC. To understand the organization of the circuit from the pIC to the CeA and

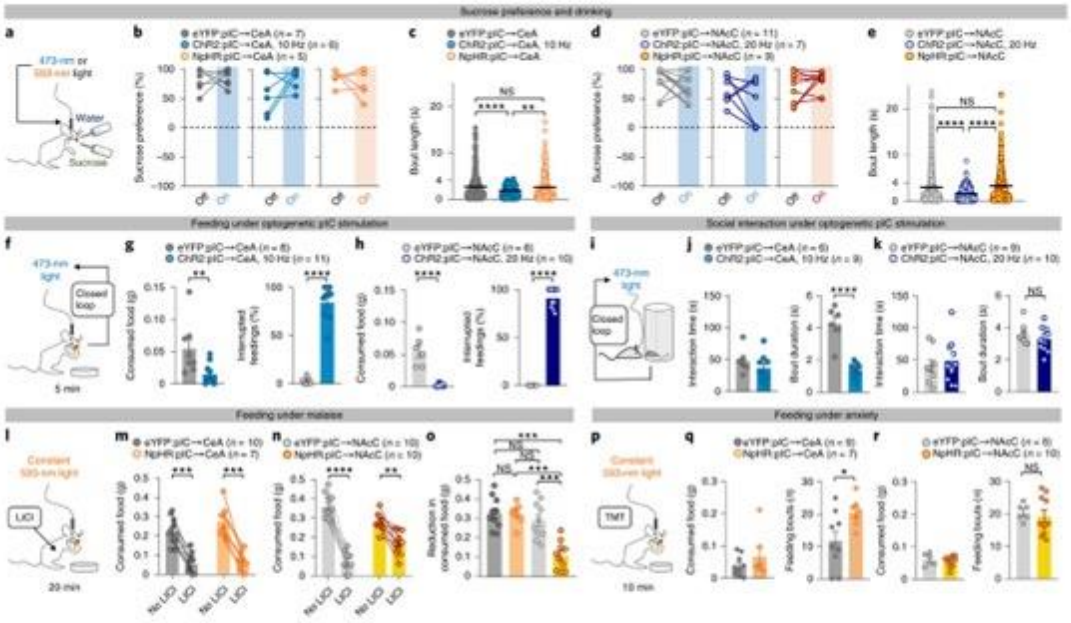


Fig. 8 | Distinct pIC outputs inhibit consumption upon the detection of homeostatic adversity or predator threat. a–e, Sucrose preference and drinking behavior upon pathway-specific optogenetic manipulations. **a**, Illustration of the open-loop sucrose preference paradigm. **b**, Sucrose preference is unaffected by pIC→CeA manipulations ($n = 7$ eYFP-, $n = 6$ ChR2- and $n = 5$ NpHR-expressing mice; two-tailed paired t-tests comparing laser-on and laser-off days: eYFP: $t = 0.2939$, d.f. = 6, $P = 0.7788$; ChR2: $t = 1.529$, d.f. = 5, $P = 0.1869$; NpHR: $t = 0.8977$, d.f. = 4, $P = 0.4201$). **c**, pIC→CeA stimulation at 10 Hz interrupts ongoing drinking. Drinking bouts are reduced in length to the 4-s interstimulus interval (ISI) ($n = 301$ bouts for eYFP, 316 bouts for ChR2 and 248 bouts for NpHR; one-way ANOVA: $F_{(2,\ 862)} = 11.25$, $P < 0.0001$; Bonferroni post hoc analysis: ChR2 versus eYFP, ****$P < 0.0001$; NpHR versus eYFP, $P > 0.9999$; ChR2 versus NpHR, **$P = 0.0012$). **d**, Sucrose preference is unaffected by pIC→NAcC manipulations ($n = 11$ eYFP-, $n = 7$ ChR2- and $n = 9$ NpHR-expressing mice; two-tailed paired t-tests comparing laser-on and laser-off days: eYFP: $t = 1.538$, d.f. = 10, $P = 0.1550$; ChR2: $t = 1.176$, d.f. = 6, $P = 0.2842$; NpHR, $t = 1.106$, d.f. = 8, $P = 0.3008$). **e**, pIC→NAcC stimulation at 20 Hz interrupts ongoing drinking. Drinking bouts are reduced in length to the 4-s ISI ($n = 700$ bouts for eYFP, 780 bouts for ChR2 and 334 bouts for NpHR; one-way ANOVA: $F_{(2,\ 1813)} = 58.26$, $P < 0.0001$; Bonferroni post hoc analysis: NpHR versus eYFP, $P = 0.2974$; ChR2 versus eYFP or NpHR, both ****$P < 0.0001$). **f**, Experimental design for closed-loop manipulations during feeding bouts. **g**, pIC→CeA stimulation in $n = 8$ eYFP- and $n = 11$ ChR2-expressing mice reduces food consumption (left; two-tailed unpaired t-test: $t = 2.973$, d.f. = 17, **$P = 0.0085$) and reliably interrupts ongoing feeding bouts (right; two-tailed unpaired t-test: $t = 13.12$, d.f. = 17, ****$P < 0.0001$). **h**, pIC→NAcC stimulation in $n = 6$ eYFP- and $n = 10$ ChR2-expressing mice strongly reduces food consumption (left; two-tailed unpaired t-test: $t = 7.271$, d.f. = 14, ****$P < 0.0001$) and reliably interrupts ongoing feeding bouts (right; two-tailed unpaired t-test: $t = 21.51$, d.f. = 14, ****$P < 0.0001$). **i**, Experimental design for closed-loop manipulations during social interactions. **j**, Although pIC→CeA stimulation in $n = 6$ eYFP- and $n = 9$ ChR2-expressing mice does not affect the overall time the mice interact (left; two-tailed unpaired t-test: $t = 1.048$, d.f. = 13, $P = 0.3138$), stimulation acutely interrupts ongoing social interactions, as demonstrated by a significant reduction in the duration of individual interaction bouts (right; two-tailed unpaired t-test: $t = 6.908$, d.f. = 13, ****$P < 0.0001$). **k**, pIC→NAcC stimulation does not affect social interactions ($n = 9$ eYFP- and $n = 10$ ChR2-expressing mice; left, total interaction time (two-tailed unpaired t-test: $t = 0.5167$, d.f. = 17, $P = 0.6120$); right, bout length (two-tailed unpaired t-test: $t = 0.8693$, d.f. = 17, $P = 0.3968$)). **l**, Experimental design for constant pathway inhibition during feeding under visceral malaise. **m**,**n**, Food consumption in hungry mice is reduced under the influence of LiCl, whether or not the pIC→NAcC or pIC→CeA pathway is inhibited. **m**, pIC→CeA groups: food consumption is reduced by LiCl under constant orange light in both eYFP-expressing (left; $n = 10$ mice; two-tailed paired t-test: $t = 5.143$, d.f. = 9, ***$P = 0.0006$) and NpHR-expressing (right; $n = 7$ mice; two-tailed paired t-test: $t = 6.844$, d.f. = 6, ***$P = 0.0005$) mice. **n**, Food consumption is also reduced under the same conditions when the pIC→NAcC pathway is manipulated (left: $n = 10$ eYFP-expressing mice; two-tailed paired t-test: $t = 9.517$, d.f. = 9, ****$P < 0.0001$; right: $n = 10$ NpHR-expressing mice; two-tailed paired t-test: $t = 4.629$, d.f. = 9, **$P = 0.0012$). **o**, In the same mice (same color code) as in **m** and **n**, however, pIC→NAcC inhibition reduces the anorexigenic effects of LiCl on food consumption (for pIC→CeA, $n = 10$ eYFP-expressing and $n = 7$ NpHR-expressing mice; for pIC→NAcC, $n = 10$ eYFP-expressing and $n = 10$ NpHR-expressing mice; one-way ANOVA: $F_{(3,\ 33)} = 16.36$, $P < 0.0001$; Bonferroni post hoc analysis demonstrated significant differences (****$P < 0.0001$) for NpHR:pIC→NAcC versus NpHR:pIC→CeA, NpHR:pIC→NAcC versus eYFP:pIC→CeA and NpHR:pIC→NAcC versus eYFP:pIC→NAcC; no differences ($P > 0.999$) were detected for NpHR:pIC→CeA versus eYFP:pIC→CeA, NpHR:pIC→CeA versus eYFP:pIC→NAcC or eYFP:pIC→NAcC versus eYFP:pIC→CeA). **p**, Experimental design for constant pathway inhibition during feeding under predator threat (synthetic fox odor, TMT). **q**, The overall food consumed under pIC→CeA inhibition ($n = 9$ eYFP-expressing and $n = 7$ NpHR-expressing mice) in the presence of TMT is unaltered (left; two-tailed unpaired t-test: $t = 1.089$, d.f. = 14, $P = 0.2944$); however, inhibited mice perform more feeding bouts than controls in the presence of TMT (right; two-tailed unpaired t-test: $t = 2.434$, d.f. = 14, *$P = 0.0289$). **r**, pIC→NAcC inhibition ($n = 6$ eYFP-expressing and $n = 10$ NpHR-expressing mice) has no effect on the amount of consumed food (left; two-tailed unpaired t-test: $t = 0.5998$, d.f. = 14, $P = 0.5582$) or on the number of feeding bouts (right; two-tailed unpaired t-test: $t = 0.4547$, d.f. = 14, $P = 0.6563$) in the presence of TMT. Bar graphs show mean ± s.e.m.

NAcC, we examined whether the same pIC neurons would send collateralizing axonal projections to both regions or whether distinct subpopulations of pIC neurons innervated these two targets. Retrograde tracings generated following injection of red and green fluorescently tagged cholera toxin subunit B (CTB) into the CeA and NAcC, respectively, of the same mice (Fig. 6a–c) showed mostly non-overlapping populations, with <6% of pIC neurons sending divergent projections to both structures (Fig. 6d–f). Interestingly, we observed increased numbers of CeA projections and decreased numbers of NAcC projectors along the anterior–posterior axis of the IC (Fig. 6f). Within the pIC, CeA projectors were more frequent (Fig. 6f). Next, we determined the distribution of pIC axons within the target regions by injecting eYFP-encoding AAVs into the pIC and determining the subregions of both structures that exhibited the densest pIC innervation (Fig. 6g–j). These regions were targeted in subsequent experiments via optic fiber implants for terminal-specific optogenetic manipulations (Figs. 7 and 8).

CeA mediates pIC top-down control of anxiety and autonomic responses. To address the functional roles of the pIC→CeA and pIC→NAcC pathways, we transduced pIC projection neurons as previously described with viruses encoding ChR2, NpHR or eYFP and placed optic fibers bilaterally above the CeA (Fig. 7a and Supplementary Fig. 13) or NAcC (Fig. 7b and Supplementary Fig. 14). We confirmed that backpropagation of terminal stimulation was limited, as behavioral effects were potently disrupted upon blocking glutamatergic transmission within the target region (Supplementary Fig. 15).

Photostimulation of pIC-derived axon terminals in the CeA increased immobility at high frequencies (20 Hz; Supplementary Fig. 16a), but did not affect locomotion at 5 or 10 Hz. In contrast, activation of the pIC→NAcC pathway had no effects on locomotion at any frequency tested (Supplementary Fig. 16e).

We next examined whether high stimulation frequencies would elicit direct behavioral reactions, as seen when stimulating pIC cell bodies. In comparison to pIC cell body stimulations, pIC→CeA stimulation at 20 Hz elicited a more restricted response pattern, including stopping, freezing and jaw movements (Fig. 7c,d). In contrast, many stimulations above the pIC terminals in the NAcC did not prompt any reaction and only occasionally elicited grooming or jaw movements (Fig. 7e,f). These data suggest that a subset of the defensive reactions observed upon pIC stimulation are mediated via glutamatergic projections to the CeA as the emotional output region.

Given the strong effects observed upon 20-Hz stimulation of pIC→CeA projections, all subsequent tests were carried out at 10 Hz. In the absence of strong effects at 20 Hz in pIC→NAcC projections, we stimulated this pathway at 20 Hz in all subsequent experiments.

We next examined whether either of the two output pathways would mediate the avoidance behavior observed upon pIC stimulation in the RTPA assay. While stimulation of the pIC→CeA pathway elicited robust aversion (Fig. 7g), as reported in previous studies[39], inhibition of either pathway (Supplementary Fig. 16b,f) and stimulation of pIC→NAcC projections (Fig. 7h) did not affect real-time place aversion.

pIC→CeA stimulation also increased respiratory rate (Fig. 7i and Supplementary Fig. 16c), whereas the same stimulation of the pIC→NAcC pathway did not affect breathing (Fig. 7j and Supplementary Fig. 16g). In line with our findings in the pIC, heart rate was not affected by projection-specific stimulations (Supplementary Fig. 16d,h).

We next determined the influence of pathway-specific activity manipulations on anxiety by using the same experimental approach as for the pIC (Fig. 2). Interestingly, manipulating the activity of the pIC→CeA pathway evoked behavioral alterations that were similar to those observed upon pIC manipulation: pIC→CeA inhibition had strong anxiolytic effects in the EPM under high-anxiety settings (Fig. 7k), but did not affect exploration of the EZM under low-anxiety settings (Fig. 7l). Inversely, pIC→CeA activation had no effects under high-anxiety settings in the EPM (Fig. 7k) but had strong anxiogenic effects in low-anxiety settings of the EZM (Fig. 7l). Notably, the same manipulations did not influence anxiety-related behaviors in the EPM or EZM when the activity of the pIC→NAcC pathway was targeted (Fig. 7m,n). These results suggest that the pIC→CeA but not the pIC→NAcC pathway mediates pIC control over anxiety-related behaviors.

pIC→NAcC projections mediate top-down inhibition of consummatory behaviors. The absence of behavioral changes upon pIC→NAcC manipulations in the context of autonomic functions, avoidance and anxiety-related behavior suggests a high specialization of top-down projections from the pIC to subcortical target regions. Because both the nucleus accumbens and CeA are known to regulate reward and consummatory behaviors[37,39,40], we carried out a sequence of experiments aimed at dissecting their specific functions in the context of appetitive behaviors.

First, we found that neither photoactivation nor photoinhibition of either IC projection pathway altered mice's preference for sucrose (Fig. 8a,b,d), nor did the manipulations affect aversion to quinine (Supplementary Fig. 17a,b). However, when either of the two pathways was stimulated while a mouse was actively drinking (water or sucrose solution), the mouse immediately interrupted ongoing drinking, as demonstrated by a significant reduction in the length of drinking bouts (Fig. 8c,e). To assess the specificity of these effects on consummatory behaviors, we next assessed feeding behavior (Fig. 8f). Activation of both projection pathways elicited immediate interruptions in feeding and an overall reduction in food consumption (Fig. 8g,h). We wondered whether the interruptions in feeding and drinking might generalize to other rewarding behaviors and tested mice while they socially interacted. Interestingly, we found interruptions of ongoing social interactions when the pIC→CeA but not the pIC→NAcC projection pathway was stimulated during interactions (Fig. 8i–k). However, the total interaction time remained unchanged upon pIC→CeA activation, which may be explained by unaltered social motivation. Indeed, the number of interaction bouts showed a trend toward increasing upon pIC→CeA stimulation (Supplementary Fig. 17e). Thus, the behavioral inhibition seemed to be more generalized upon pIC→CeA stimulation, in line with an anxiety signal that may transiently stop ongoing behaviors.

To address what the physiological roles of each of these projection pathways might be, we used optogenetic inhibition experiments to assess the necessity of activity in these pathways when feeding behavior was naturally interrupted or reduced.

The existing literature and our own findings strongly suggest a role for the pIC in representing aversive bodily and emotional states, such as malaise and anxiety[32]. Because both of these states powerfully inhibit feeding, we next tested the necessity of the two pIC-emergent pathways in mediating the feeding inhibition experienced under these aversive states.

We first focused on malaise-induced anorexia in hungry mice (Fig. 8l). All mice exhibited significant reductions in feeding in comparison to baseline when LiCl was administered before feeding (Fig. 8m–o). However, mice in which the pIC→NAcC pathway was silenced exhibited a smaller reduction in feeding than control mice or mice in which the pIC→CeA pathway was inhibited (Fig. 8o), suggesting that pIC→NAcC projections transmit information about the aversive state of malaise to the NAcC to elicit feeding inhibition. In a second set of experiments, we addressed whether anxiety, as another potent inhibitor of consumption, might also be transmitted through either of the two pIC pathways. When hungry mice were

allowed to feed in the presence of anxiety-evoking synthetic fox odor (TMT), all mice exhibited a strong reduction in the number of feeding bouts (Fig. 8p–r and Supplementary Fig. 17d). Interestingly, mice in which the pIC→CeA but not the pIC→NAcC pathway was inhibited exhibited a significant increase in the number of feeding bouts during anxiety-induced anorexia (Fig. 8q,r). These data suggest that pIC→CeA silencing helped mice to overcome predator-odor-mediated avoidance of the food source, without affecting food consumption itself.

Notably, inhibiting the pIC→NAcC or pIC→CeA pathway in hungry but otherwise naive mice did not alter the amount of food the mice consumed, suggesting that the pIC specifically transmits the aversive states of malaise, disgust or fear to its subcortical targets (Supplementary Fig. 17c).

Taken together, our results demonstrate highly specialized functions of two separate top-down projection pathways emerging from the pIC. While the pIC→CeA pathway mediates anxiety-related behaviors and has a generalized role in suppressing appetitive behaviors upon the detection of threat, the pIC→NAcC pathway contributes to the inhibition of ongoing consummatory behavior upon the detection of homeostatic adversity, such as malaise.

Discussion

Internal states powerfully modulate behavior. Our study shows how a subregion of the IC, the posterior visceral or 'interoceptive' insula, represents aversive emotional and bodily states and exerts top-down regulation of ongoing behavior. We comprehensively map the multimodal convergence of pIC inputs and the top-down projections to subcortical targets mediating emotional and motivated functions.

While human imaging studies have suggested a role for the IC in processing negative emotions, such as anxiety or low mood in depression[39,41–43], our study provides a comprehensive description of neuronal circuit mechanisms that underlie the detection and processing of diverse aversive states within the pIC by using mice as a model organism. Notably, our findings underline the extent to which the pIC should be regarded as a multimodal hub, processing stimuli and states of very different temporal dynamics and origins: sensory, emotional and bodily.

Although we observed that subsets of pIC neurons were readily activated by acute sensory stimuli, such as painful tail shocks, silencing of the pIC did not affect acute reactions to pain or pain thresholds, similarly to findings reported in a recent study[23]. However, the same study[23] reported that inhibition of the pIC when foot shocks were applied during fear conditioning reduced acute freezing behavior and weakened the formation of threat memories during later auditory fear recall. In contrast to these findings, we were unable to detect any changes in acute freezing behavior or in the formation of contextual or cued fear memories upon constant pIC inhibition during aversive foot shocks. Instead, we found that pIC silencing impaired the emergence of sustained states of anxiety. Furthermore, we found that persistent anxiety states were represented in large fractions of pIC neurons, and our results show that both diverse acute sensory stimuli and sustained aversive internal states are processed within single neurons of the pIC. Interestingly, our own work and the aforementioned study describe different aspects of fear and anxiety regulation as being mediated by strong projections from the pIC to the lateral and central amygdalae. Thus, while much remains to be understood about the aversive coding properties of the pIC, it is clearly a crucial brain region in processing and modulating aversive emotions.

Our study further dissects the contribution of distinct output pathways mediating the behavioral consequences of multiple internal aversive states, such as reductions in exploration and feeding upon bodily and affective adversity, thus highlighting that multiple facets of aversive state processing are mediated at the level of segregated subcortical pIC projections.

Because our present findings uncover negative-valence processing in the pIC, it is interesting to note that previous studies advocated for the existence of a valence map for neurons in the gustatory IC, which, when separately activated, drive aversion or attraction[44,45]. While our pIC region partially overlaps the 'bitter' portion of the gustatory cortex[44], our anatomical, behavioral and imaging data provide evidence for a much more general role in the processing of aversive stimuli and states within the pIC. As previous studies have identified appetitive regions at more anterior sites in the IC[44,46], an important open question is whether positive internal states are represented within a different topographical region of the IC and how such a circuit would interact with the pIC.

While the neural substrates of aversive states are likely to be found at the level of extended neuronal networks rather than isolated brain regions[47,48], we here provide a mechanistic basis for behavioral top-down regulation caused by the detection of aversive emotional and bodily states in the pIC. Notably, the IC has been suggested to be a neuroanatomical hub underlying various forms of severe neuropsychiatric conditions in humans[14,19–21,42,49]. Our findings thus provide an entry point for mechanistic investigation of the insular contribution to a variety of neuropsychiatric conditions, including but not limited to anxiety disorders, major depression, addiction and eating disorders.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41593-019-0469-1.

Received: 6 February 2019; Accepted: 19 July 2019;
Published online: 27 August 2019

References

1. Lovett-Barron, M. et al. Ancestral circuits for the coordinated modulation of brain state. *Cell* **171**, 1411–1423 (2017).
2. Zelikowsky, M. et al. The neuropeptide Tac2 controls a distributed brain state induced by chronic social isolation stress. *Cell* **173**, 1265–1279 (2018).
3. Contreras, M., Ceric, F. & Torrealba, F. Inactivation of the interoceptive insula disrupts drug craving and malaise induced by lithium. *Science* **318**, 655–658 (2007).
4. Kurth, F., Zilles, K., Fox, P. T., Laird, A. R. & Eickhoff, S. B. A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Struct. Funct.* **214**, 519–534 (2010).
5. Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A. & Dolan, R. J. Neural systems supporting interoceptive awareness. *Nat. Neurosci.* **7**, 189–195 (2004).
6. Craig, A. D. Interoception: the sense of the physiological condition of the body. *Curr. Opin. Neurobiol.* **13**, 500–505 (2003).
7. Simmons, W. K. et al. Keeping the body in mind: insula functional organization and functional connectivity integrate interoceptive, exteroceptive, and emotional awareness. *Hum. Brain Mapp.* **34**, 2944–2958 (2013).
8. Allen, G. V., Saper, C. B., Hurley, K. M. & Cechetto, D. F. Organization of visceral and limbic connections in the insular cortex of the rat. *J. Comp. Neurol.* **311**, 1–16 (1991).
9. Cechetto, D. F. & Saper, C. B. Evidence for a viscerotopic sensory representation in the cortex and thalamus in the rat. *J. Comp. Neurol.* **262**, 27–45 (1987).
10. Craig, A. D. How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* **3**, 655–666 (2002).
11. Tan, L. L. et al. A pathway from midcingulate cortex to posterior insula gates nociceptive hypersensitivity. *Nat. Neurosci.* **20**, 1591–1601 (2017).
12. Gogolla, N., Takesian, A. E., Feng, G., Fagiolini, M. & Hensch, T. K. Sensory integration in mouse insular cortex reflects GABA circuit maturation. *Neuron* **83**, 894–905 (2014).
13. Livneh, Y. et al. Homeostatic circuits selectively gate food cue responses in insular cortex. *Nature* **546**, 611–616 (2017).
14. Singer, T., Critchley, H. D. & Preuschoff, K. A common role of insula in feelings, empathy and uncertainty. *Trends Cogn. Sci.* **13**, 334–340 (2009).
15. Etkin, A., Büchel, C. & Gross, J. J. The neural bases of emotion regulation. *Nat. Rev. Neurosci.* **16**, 693–700 (2015).
16. Etkin, A. & Wager, T. D. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am. J. Psychiatry* **164**, 1476–1488 (2007).

17. Grupe, D. W. & Nitschke, J. B. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat. Rev. Neurosci.* **14**, 488–501 (2013).
18. Nagai, M., Kishi, K. & Kato, S. Insular cortex and neuropsychiatric disorders: a review of recent literature. *Eur. Psychiatry* **22**, 387–394 (2007).
19. Downar, J., Blumberger, D. M. & Daskalakis, Z. J. The neural crossroads of psychiatric illness: an emerging target for brain stimulation. *Trends Cogn. Sci.* **20**, 107–120 (2016).
20. Goodkind, M. et al. Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* **72**, 305–315 (2015).
21. Namkung, H., Kim, S.-H. & Sawa, A. The insula: an underestimated brain area in clinical neuroscience, psychiatry, and neurology. *Trends Neurosci.* **40**, 200–207 (2017).
22. Casanova, J. P. et al. A role for the interoceptive insular cortex in the consolidation of learned fear. *Behav. Brain Res.* **296**, 70–77 (2016).
23. Berret, E. et al. Insular cortex processes aversive somatosensory information and is crucial for threat learning. *Science* **364**, eaaw0474 (2019).
24. Contreras, M. et al. A role for the insular cortex in long-term memory for context-evoked drug craving in rats. *Neuropsychopharmacology* **37**, 2101–2108 (2012).
25. Foilb, A. R., Flyer-Adams, J. G., Maier, S. F. & Christianson, J. P. Posterior insular cortex is necessary for conditioned inhibition of fear. *Neurobiol. Learn. Mem.* **134**, 317–327 (2016).
26. Christianson, J. P. et al. Safety signals mitigate the consequences of uncontrollable stress via a circuit involving the sensory insular cortex and bed nucleus of the stria terminalis. *Biol. Psychiatry* **70**, 458–464 (2011).
27. Rogers-Carter, M. M. et al. Insular cortex mediates approach and avoidance responses to social affective stimuli. *Nat. Neurosci.* **21**, 404–414 (2018).
28. Hanamori, T., Kunitake, T., Kato, K. & Kannan, H. Responses of neurons in the insular cortex to gustatory, visceral, and nociceptive stimuli in rats. *J. Neurophysiol.* **79**, 2535–2545 (1998).
29. Oppenheimer, S., Gelb, A., Girvin, J. & Hachinski, V. Cardiovascular effects of human insular cortex stimulation. *Neurology* **42**, 1727–1732 (1992).
30. Yasui, Y., Breder, C. D., Safer, C. B. & Cechetto, D. F. Autonomic responses and efferent pathways from the insular cortex in the rat. *J. Comp. Neurol.* **303**, 355–374 (1991).
31. Chen, X., Gabitto, M., Peng, Y., Ryba, N. J. P. & Zuker, C. S. A gustotopic map of taste qualities in the mammalian brain. *Science* **333**, 1262–1266 (2011).
32. Pastuskovas, C. V., Cassell, M. D., Johnson, A. K. & Thunhorst, R. L. Increased cellular activity in rat insular cortex after water and salt ingestion induced by fluid depletion. *Am. J. Physiol. Integr. Comp. Physiol.* **284**, R1119–R1125 (2003).
33. Reimer, J. et al. Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron* **84**, 355–362 (2014).
34. Wickersham, I. R., Finke, S., Conzelmann, K.-K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nat. Methods* **4**, 47–49 (2007).
35. Shi, C. J. & Cassell, M. D. Cortical, thalamic, and amygdaloid connections of the anterior and posterior insular cortices. *J. Comp. Neurol.* **399**, 440–468 (1998).
36. Mufson, E. J., Mesulam, M.-M. & Pandya, D. N. Insular interconnections with the amygdala in the rhesus monkey. *Neuroscience* **6**, 1231–1248 (1981).
37. Namburi, P., Al-Hasani, R., Calhoon, G. G., Bruchas, M. R. & Tye, K. M. Architectural representation of valence in the limbic system. *Neuropsychopharmacology* **41**, 1697–1715 (2015).
38. Floresco, S. B. The nucleus accumbens: an interface between cognition, emotion, and action. *Annu. Rev. Psychol.* **66**, 25–52 (2015).
39. Fadok, J. P., Markovic, M., Tovote, P. & Lüthi, A. New perspectives on central amygdala function. *Curr. Opin. Neurobiol.* **49**, 141–147 (2018).
40. Schiff, H. et al. An insula–central amygdala circuit for guiding tastant-reinforced choice behavior. *J. Neurosci.* **38**, 1418–1429 (2018).
41. Avery, J. A. et al. Major depressive disorder is associated with abnormal interoceptive activity and functional connectivity in the insula. *Biol. Psychiatry* **76**, 258–266 (2014).
42. Paulus, M. P. & Stein, M. B. Interoception in anxiety and depression. *Brain Struct. Funct.* **214**, 451–463 (2010).
43. Fitzgerald, P. B., Laird, A. R., Maller, J. & Daskalakis, Z. J. A meta-analytic study of changes in brain activation in depression. *Hum. Brain Mapp.* **29**, 683–695 (2008).
44. Peng, Y. et al. Sweet and bitter taste in the brain of awake behaving animals. *Nature* **527**, 512–515 (2015).
45. Wang, L. et al. The coding of valence and identity in the mammalian taste system. *Nature* **558**, 127–131 (2018).
46. Castro, D. C. & Berridge, K. C. Opioid and orexin hedonic hotspots in rat orbitofrontal cortex and insula. *Proc. Natl Acad. Sci. USA* **114**, E9125–E9134 (2017).
47. Damasio, A. & Carvalho, G. B. The nature of feelings: evolutionary and neurobiological origins. *Nat. Rev. Neurosci.* **14**, 143–152 (2013).
48. LeDoux, J. Rethinking the emotional brain. *Neuron* **73**, 653–676 (2012).
49. Naqvi, N. H. & Bechara, A. The hidden island of addiction: the insula. *Trends Neurosci.* **32**, 56–67 (2009).

Acknowledgements

We thank A. Ghanem (Ludwig Maximilians University) for producing modified rabies viruses; W. Denk, I. Grundwald-Kadow, R. Klein and R. Portugues for critical reading of earlier versions of this manuscript; K. Deisseroth (Stanford University) for optogenetic and Cre-dependent AAV constructs and the UNC Vector Core for viral packaging; F. Lyonnaz for managing the animal colony; and C. Weiand for technical assistance. This study was supported by the Max Planck Society, the Deutsche Forschungsgemeinschaft (SPP1665 to K.-K.C., D.A.G. and N.G.), funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-2017-STG, grant agreement 758448 to N.G.), a German–Israeli Foundation grant (to N.G. and N.R.V., grant I-1301-418.13/2015) and the ANR-DFG project SafeNet (project no. 391081777 to N.G. and A.S.K.).

Author contributions

D.A.G. and N.G. designed the study and analyzed data. D.A.G., A.S.K., M.J. and A.M. performed optogenetic surgeries, behavior experiments and analyses. N.D. performed all two-photon imaging experiments and analyses, and helped with physiological recordings. R.R.C. and A.S.K. performed fiber photometry recordings. R.R.C., D.A.G. and A.S.K. performed photometry analyses. A.S.K. performed optrode recordings. N.R.V. assisted with behavior analysis and provided custom-written code. T.D.B. and A.P. helped with the histology. K.-K.C. provided rabies virus and shared expertise in monosynaptic tracing. D.A.G. performed all tracing experiments. D.A.G. and T.N.G. analyzed tracing experiments and performed immunohistochemistry. N.G. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-019-0469-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to N.G.

Peer review information: *Nature Neuroscience* thanks Wulf Haubensak and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Mice. Mice from 2–6 months of age were used in accordance with the regulations from the government of Upper Bavaria. Male C57BL/6NRj mice were used for optogenetic experiments. Retrograde rabies tracings and axonal AAV tracings employed male or female *Camk2a*-Cre (B6.Cg-Tg(Camk2a-cre)T29–1Stl/J]) mice. Animals used in behavioral experiments were housed in pairs of two. All mice were kept on an inverse 12-h light/12-h dark cycle (lights off at 11:00). Mice were provided with ad libitum access to standard chow and water. For the sucrose preference test and quinine aversion test, mice were deprived of water for 24 h. For the feeding interruption test, the TMT test and the feeding under malaise test, mice were acutely food restricted for 24 h before testing.

Viral constructs. For in vivo optogenetic experiments, axonal AAV tracings and retrograde rabies tracings, the following constructs were obtained from the UNC Vector Core (Gene Therapy Center, University of North Carolina at Chapel Hill): AAV2/5-*Camk2a*-hChR2(H134R)-eYFP (6.2×10^{12} particles per ml), AAV2/5-*Camk2a*-eNpHR3.0-eYFP (5.2×10^{12} particles per ml), AAV2/5-*Camk2a*-eYFP (4.3×10^{12} viral genomes (vg) ml^{-1}), AAV2/5-EF1α-DIO-eYFP (5.6×10^{12} $vg\,ml^{-1}$) and helper viruses for retrograde rabies tracings: AAV2/8-EF1α-FLEX-TVAmCherry (4.2×10^{12} $vg\,ml^{-1}$) and AAV2/8-CA-FLEX-RG (2.5×10^{12} $vg\,ml^{-1}$). For retrograde rabies tracings, we used glycoprotein (G)-deleted EnvA-pseudotyped rabies-GFP (RABVΔG-GFP(EnvA), 3×10^8 focus-forming units (FFU) ml^{-1}), which was kindly provided by K.K. Conzelmann (Max von Pettenkofer Institute and Gene Center, Ludwig Maximilians University). For photometry, we used AAV9.*Camk2a*.GCaMP6s.WPRE.SV40 ($\geq1\times10^{13}$ $vg\,ml^{-1}$) obtained from Addgene (viral preparation 107790-AAV9). For two-photon imaging, we used an AAV9.hSyn.GCaMP6s.WPRE.SV40 vector ($\geq1\times10^{13}$ $vg\,ml^{-1}$) obtained from the University of Pennsylvania Vector Core. For DREADD experiments, we used AAV8-hSyn-DIO-hM3D(G_q)-mCherry (4×10^{12} $vg\,ml^{-1}$; Addgene, 44361), AAV8-*Camk2a*-hM4D(G_i)-mCherry (2.64×10^{12} GC ml^{-1}; Addgene, 50477) and AAV2/5-*Camk2a*-eYFP (4.3×10^{12} $vg\,ml^{-1}$; UNC Vector Core).

Surgeries. We injected mice with metamizol (200 mg per kg, subcutaneously) for perioperative analgesia and carprofen (5 mg per kg, subcutaneously once daily for 3 d) for postoperative pain care. Anesthesia (isoflurane) was initiated at 5% and maintained at 1–2.5% throughout surgery. Mice were secured in a stereotaxic frame (Stoelting) and placed on a heating pad (37 °C), and eye ointment (Bepanthen, Bayer) was applied. We performed two skull trepanations for manipulation of pIC cell bodies and four trepanations for projection-specific manipulation of axon terminals. For viral injections, we used pulled-glass pipettes attached to a microliter syringe (5-µl model 75 RN, Hamilton) via a glass needle compression fitting (55750-01, Hamilton), mounted on a syringe pump controlled by a microcontroller (UMP3 with a Micro4, WPI).

For in vivo optogenetic experiments, mice were bilaterally injected with 150 nl of virus at a rate of 80 nl min^{-1} in the pIC (distances from bregma: AP: −0.45 mm, ML: ±4.05 mm, DV: −4.05 mm). Custom-made optic fibers (200-µm core, 0.22 NA, 1.25-mm zirconia ferrule from Thorlabs) were implanted 0.5 mm above the injection sites.

For fiber photometry, 200 nl of AAV9-*Camk2a*-GCaMP6s virus was injected unilaterally into the pIC (coordinates: AP: −0.5 mm, ML: ±4 mm, DV: −4 mm) and a fiber optic cannula (custom-made optic fiber of 200 µm, 0.48 NA, glued to zirconia ferrules (2.5 mm)) was inserted 0.2 mm above the injection site. Cannulae were secured to the skull with acrylic glue.

For projection-specific optogenetic manipulations, mice were bilaterally injected with 150 nl of virus into the pIC as described above, but optic fibers were bilaterally implanted over the CeA (AP: −1.3 mm, ML: ±2.8 mm, DV: −4.3 mm) or above the NAcC (AP: +0.9 mm, ML: ±2.05 mm, DV: −3.9 mm; 10° outward angle in the coronal plane).

For the backpropagation experiments, we bilaterally injected 150 nl of AAV-*Camk2a*-hChR2(H134R)-eYFP into the pIC as described above and then implanted opto-fluid cannulae (iOFC_M3_320/430 and DiOFC_L_ZF_2.8_320/430, Doric Lenses), which are commercial guide cannulas through which optic fibers and fluid-injectors can be sequentially inserted, over the CeA (AP: −1.3 mm, ML: ±2.8 mm, DV: −4.0 mm; with an optic insert at DV: −4.3 mm or an infusion insert at DV: −4.7 mm) or NAcC (AP: +0.9 mm, ML: ±1.4 mm, DV: −4.0 mm; with an optic insert at DV: −4.2 mm or an infusion insert at DV: −4.6 mm).

We sealed the space between the fiber and the trepanation with bone wax to protect the underlying brain tissue from the cyanoacrylate glue (Ultra Gel, Pattex) that secured the optic fibers to the skull. Additionally, we added a layer of black dental cement to further secure the optic fibers and to reduce light emission from the skull.

For two-photon calcium imaging, we injected a total of 450 nl of AAV9-hSyn-GCaMP6s into the pIC, with 150 nl each injected at three dorso-ventral positions (distances from the bregma: AP: −0.45 mm, ML: ±4.05 mm, DV: −3.7, −3.5 and −4.05 mm).

For monosynaptic retrograde rabies tracings, mice were first unilaterally injected with 100–150 nl of a 6:1 (RG:TVA) mixture of helper viruses in the pIC. The trepanation was sealed with bone wax, and the skin was sutured. After 3–4 weeks, 350 nl of RABVΔG-GFP(EnvA) was injected at the same coordinates. Mice were killed 7 d after infusion of the rabies virus.

For axonal AAV tracings, we unilaterally injected 80–100 nl of AAV2/5-EF1α-DIO-eYFP into the pIC of *Camk2a*-Cre mice. Four weeks after the injections, mice were killed.

For CTB double labeling, C57BL/6NRj mice were unilaterally injected with 150 nl of a 0.5% CTB-555 solution (in PBS) in the CeA (AP: −1.3 mm, ML: ±2.8 mm, DV: −4.7 mm) and 300 nl of a 0.5% CTB-488 solution in the NAcC (AP: +0.9 mm, ML: 1.35 mm, DV: −4.7 mm). Mice were killed 7 d after CTB injections.

For cranial window surgery for two-photon calcium imaging, 2 weeks after AAV-GCaMP6s injection into the pIC, C57BL/6NRj mice (12–18 weeks old) were anesthetized with a mixture of midazolam, medetomidin and fentanyl (5 mg per kg, 0.5 mg per kg and 0.05 mg per kg, respectively) and injected with dexamethasone (subcutaneously, 0.2 mg per kg). Lidocaine was applied topically at the incision site. A head post was secured to the skull with C&B Super-Bond (Sun Medical). Tissue above the left IC was removed, and the skull was cleaned. A craniotomy was created by using a 3-mm biopsy punch (Miltex) and positioned with the rhinal vein as the ventral limit and the medial cerebral artery slightly anterior to the center of the window, as seen through the skull. Because the cortical surface is curved, the brain was protected from being flattened by applying a small amount of transparent polymer, as previously described[34]. A 3-mm coverslip was gently positioned on top of the polymer and secured with cyanoacrylate glue. The remaining exposed skull was covered with C&B Super-Bond.

For DREADD experiments, C57BL/6NRj or *Camk2a*-Cre mice were bilaterally injected with 200 nl of virus per hemisphere in the pIC as described above.

Behavioral experiments. All mice were handled by the experimenters and habituated to tethering of the fiber optic patch cords for at least 3 d. All behavioral experiments were performed during the dark phase of the light cycle between 12:00 and 20:30.

Stimulation-evoked behaviors. Optogenetics. Mice were tethered to optic patch cords and placed in a behavioral arena (20 × 50 × 30 cm^3) with transparent plexiglass walls and recorded with a CMOS camera at 100 f.p.s. (DMK 33UP1300, Imaging Source). Laser stimulation was manually triggered by the experimenter at random intervals or when behavioral reactions had subsided during the 5-min test (473 nm, 1 s, 20 Hz, 5-ms pulse width, 3–5 mW, minimum ISI = 4 s). The experiment was terminated prematurely if we observed severe reactions like jumping or crouching. Stimulation-evoked behaviors were manually scored.

DREADDs. Mice bilaterally expressing the excitatory DREADD receptor hM3Dq or eYFP in the pIC were injected with clozapine-*N*-oxide (CNO; 10 mg per kg, intraperitoneally), and after 40 min they were placed into a circular open-field arena. The mice were video recorded for 15 min. Stimulation-induced behaviors were scored manually by a human observer inspecting the videos post hoc and with blinding to the underlying condition.

Real-time place avoidance assays. The RTPA assay was performed on three consecutive days in custom-built chambers (50 × 40 × 25 cm^3, transparent plexiglass, distinct metal floor grids, distinct visual patterns presented in chambers). We tracked the mice with a webcam (c930e, Logitech) that was interfaced with custom-written MATLAB (Mathworks) software. The first day, mice were tracked for 20 min to establish their preferred chamber. The following day, the preferred chamber became the 'stimulated chamber'. The trial started in the non-stimulated chamber and lasted 20 min. Crossing into the stimulated chamber resulted in delivery of light of 473 nm (at either 10 Hz, 3–5 mW (pIC-ChR2 and pIC→CeA) or 20 Hz, 7–10 mW (pIC→NAcC); all with a 5-ms pulse width and a paradigm in which the laser was on for 1 s followed by a 4-s ISI). For NpHR-expressing mice, we applied constant 594-nm illumination at 12 mW. Exiting the stimulated chamber immediately switched off the laser. On day 3, we tested whether the mice displayed an association with the stimulation chamber (no laser). Data were analyzed with custom-written MATLAB scripts.

Real-time control of pIC activity during elevated plus maze exploration in a high-anxiety setting. The custom-built EPM had the following dimensions: 5-cm-wide arms, each of which was 35-cm long; a 5-mm ridge on the open arms to prevent falls; and closed walls 15 cm high; the maze was elevated 60 cm from the floor. EPM testing was performed for 10 min in a room to which the mice had never been exposed. The mice were minimally handled (3 d with tethering to optic fibers). We illuminated the center of the EPM with 450 lux. The environment was noisy owing to a cooling fan. We used ANYmaze (Stoelting) to track mice in real time, analyze data and trigger optogenetic manipulation on the basis of the location of the mice. Guided by our fiber photometry data (Fig. 2a,b), we inhibited the pIC or its projections to the CeA and NAcC when a mouse was located in the closed arms of the EPM, but not in the open arms. In contrast, we optogenetically stimulated the pIC or its projections to the CeA and NAcC when a mouse was in the open arms of the EPM, but not in the closed arms. The stimulation parameters were set to 10 Hz, 3–5 mW (pIC-ChR2 and pIC→CeA) or 20 Hz, 7–10 mW (pIC→NAcC), all with a 5-ms pulse width and a paradigm in which the laser was on for 1 s followed by a 4-s ISI). NpHR-expressing mice were subjected to constant 594-nm illumination at 12 mW.

Real-time control of pIC activity during elevated zero maze exploration in a low-anxiety environment. The custom-built EZM had the following dimensions: a diameter of 55 cm; 5-cm-wide arms; a 5-mm ridge on the open arms to prevent falls; and 20-cm-high closed walls; the maze was elevated 60 cm above the floor. To create a low-anxiety setting, we performed the task on well-handled mice (handled for ~6–7 d) in a quiet and dimly lit room (6 lux on the open arms) to which the mice were habituated. We used ANYmaze to track mice in real time, analyze data and trigger optogenetic manipulation on the basis of the location of the mice, as described above for real-time optogenetic control in the EPM.

Optogenetic stimulation followed by EPM. ChR2:pIC mice or eYFP:pIC controls were optogenetically stimulated for 50 s in an open-field arena at 20 Hz, 3–5 mW with a 5-ms pulse width. After 1 min of recovery, mice were disconnected from the optical patch cord and subsequently tested for 8 min on the EPM. We placed mice in the center zone of the EPM facing a closed arm. EPM behavior was analyzed with ANYmaze.

Attenuation of sustained anxiety. To induce an anxiety state, mice were placed in behavior boxes (Ugo Basile) consisting of a metal grid floor and plexiglass walls with distinct visual cues. The floor and walls were cleaned with 80% ethanol before and after each session. Freezing behavior was scored with ANYmaze. Mice were considered to be frozen if no movement except breathing was detected for at least 2 s. Mice were tethered to optic patch cords and placed in the fear conditioning box, where they were administered five unsignaled shocks (1 s, 0.4 mA) through the metal grid floor over a 20-min time period with an ISI of 160–290 s. Optogenetic inhibition (594 nm, 12 mW) was delivered over the entire 20 min. Mice were then detached from the optic patch cords and placed in the center zone of the EPM facing the closed arms and tested for 8 min. For testing of contextual fear memory recall, mice were placed in the same behavior boxes for 15 min 7 d after fear conditioning and freezing behavior was analyzed over the whole session.

Ramping hot plate test. The temperature of the plate started at 32 °C and was ramped up by 1 °C every 4 s. The test was stopped and the temperature was recorded by an experimenter blinded to experimental conditions as soon as mice licked their paws. NpHR:pIC mice were tethered to optic patch cords and habituated to the hot plate apparatus (Hot Cold Plate Analgesia Meter for Mice and Rats, Campden Instruments) for a maximum of 5 min. Each mouse was tested a total of 12 times (6 times with the light on and 6 times with the light off) on two separate days. Values are presented as the average of these 12 repeats.

Sucrose preference test and quinine avoidance test. Mice deprived of water for 24 h were trained for 30 min daily over 5 d to consume either a sucrose solution or only water, from two sipper tubes presented in a custom-built chamber ($30 \times 50 \times 30\,cm^3$), until they reached a stable preference for sucrose. Because comparable sucrose preference assays used sucrose concentrations of 1–10% (refs. [36,37]), we chose a medium sucrose concentration of 5%. The sucrose preference index was calculated as (sucrose licks − water licks)/(sucrose licks + water licks). The locations of the sucrose solution and water alternated daily, and licks were recorded via a custom-built lickometer that was interfaced with MATLAB. On test days, we delivered an open-loop stimulation pattern in which the laser was on for 1 s followed by a 4-s ISI for 20 min. pIC→CeA mice were tested at 10 Hz on separate days, and pIC→NAcC mice were tested at 20 Hz. For NpHR-expressing mice, we delivered constant illumination (12 mW) for 20 min. For the quinine avoidance test, we first established stable avoidance of quinine (0.5 mM in water) for three consecutive days. Then, on test days, we performed optogenetic manipulations as described above. Quinine preference index was calculated as (quinine licks − water licks)/(quinine licks + water licks). Drinking behavior was analyzed with custom-written Python code. For analysis of bout length, we grouped licks into a bout if the next interval between licks was greater than 0.3 s.

Feeding under closed-loop optogenetic stimulation. We habituated acutely food-deprived (24-h) mice to the behavioral chamber ($20 \times 50 \times 30\,cm^3$) for 5 min. Then, a standard chow pellet was secured to the floor with double-sided tape. We observed behavior and manually triggered laser stimulation once mice started nibbling on the food pellet. This led to laser stimulation for 1 s at 473 nm, 10 or 20 Hz, and 7–10 mW with a 5-ms pulse width, as measured at the tip of the fiber. The experimenter stimulated the mice at every attempt to feed (no 4-s ISI). The trial was recorded for 5 min at 100 f.p.s. and manually scored with ANYmaze. A successful interruption was scored if laser stimulation could stop chewing or eating for at least 1 s. Food intake was measured by weighing the food pellets and crumbles with precision fine scales before and after the trial.

Feeding under malaise. First, we tested whether inhibition of the pIC→CeA or pIC→NAcC pathway for 20 min (594 nm, 10–12 mW) influenced feeding in acutely food-deprived mice. After a day of refeeding, mice were again food deprived for 24 h and were intraperitoneally injected with a LiCl solution (0.3 M in 0.9% saline; 1.2% of body weight (for example, 350 µl for a 30-g mouse)) for 30 min before the test. Terminals of the pIC→CeA or pIC→NAcC pathway were inhibited for 20 min (594 nm, 10–12 mW, as measured at the tip of the fiber) in the presence of a standard chow pellet. The weights of the pellet before and after the test were measured with precision fine scales.

Social approach test. Mice were habituated for 5 min to a custom-built chamber ($30 \times 50 \times 30\,cm^3$) containing an empty cylindrical cage. After habituation, male juvenile C57BL/6NRj mice were placed inside the cylinder cage for 5 min. Every social interaction resulted in 1-s laser stimulation (parameters as above, no ISI) manually triggered by an experimenter. Close-up videos of the interactions were manually scored with ANYmaze.

Feeding under anxiety. First, we tested whether inhibition of the pIC in the absence of TMT would affect feeding in food-deprived mice. Mice were tethered to optic patch cords and habituated to an empty cage for 5 min, after which a standard food pellet and a 2.5-cm plastic Petri dish with saline-soaked filter paper were placed on one side of the cage. We constantly illuminated the pIC with 594-nm light (12 mW) for 10 min and recorded behavior from the top and the side. Following this control experiment, we repeated the test, but this time soaked the filter paper with 7 µl of a 90% 2,3,5-trimethyl-3-thiazoline (TMT) solution (BioSRQ). Again, the pIC was inhibited by delivering 594-nm light as described above. After each test run, we ventilated the behavioral room for 5–10 min with a rotating fan and opened the lab doors. Videos from the side were manually scored for flat-back approaches, rearing and grooming events by an observer blinded to the experimental conditions.

Frequency-dependent locomotion response. We connected mice to optic patch cords and placed them in an open arena ($50 \times 25 \times 25\,cm^3$). After a 1-min acclimatization period, mice received optogenetic stimulation at 5, 10 or 20 Hz (473 nm, 1 s on/4 s off, 3–5 mW) for 30 s with a 2-min break when switching the stimulation frequency. Locomotion parameters were analyzed with ANYmaze.

Backpropagation control experiments. To test whether the behavioral effects induced by stimulation of pIC terminals over the CeA and NAcC were indeed mediated by monosynaptic glutamatergic inputs from the pIC to the CeA or NAcC rather than by axons of passage or antidromic activation of pIC somata, we performed the following control experiment adapted from a previous report[33].

We bilaterally implanted opto-fluid cannulae over either the CeA or NAcC in mice expressing ChR2 for 6 weeks in the pIC (for information on coordinates and cannulae, see 'Surgeries'). First, we tested whether interruption of ongoing feeding upon stimulation of the CeA or NAcC pathway could be reproduced. We guided the optic inserts (Doric Lenses; CeA, OI_iOFC-M3_200/240_0.22_FLT_4.3; NAcC, OI_DiOFC-L-ZF_200/240_0.22_FLT_4.2) through the opto-fluid cannulae, connected mice to optic patch cords and performed feeding under the closed-loop optogenetic stimulation paradigm, as described above. Every attempt to feed led to 1-s stimulation (473 nm, 5 ms, 3–5 mW, 10 Hz (CeA) or 20 Hz (NAcC), no ISI). To assess effects within the same individuals, after 24 h of refeeding, the mice were again food deprived for 24 h. We then repeated feeding under closed-loop optogenetic stimulation. This time, 1–1.5 h before testing, mice were briefly anesthetized and 300 nl of the AMPA receptor antagonist NBQX (NBQX disodium salt, Enzo Life Sciences; 5 mg ml^{-1} in sterile 0.9% saline) was infused through the fluid inserts (Doric Lenses; CeA, FI_iOFC-M3_100/170_4.7; NAcC, FI_DiOFC-L-ZF_100/170_4.6). For infusions, we connected the fluid inserts to a 5-µl microsyringe (model 75 RN, Hamilton) with PE tubing and inserted them into the implanted guide cannulae. The flow rate (150 nl min^{-1}) was controlled by a syringe pump (UMP3 pump with a sys-Micro4 controller, WPI).

Anxiety-like behavior under DREADD-mediated pIC manipulations. Six weeks after bilateral injection of virus (hM4Di, hM3Dq or eYFP control) into the pIC, mice were habituated to being injected intraperitoneally for three consecutive days with 300 µl of 0.9% saline solution. On the day of the experiment, 10 mg per kg CNO solution (in sterile saline with a final DMSO concentration of 2.5%; Sigma, C0832) was injected intraperitoneally 30–40 min before testing. Mice were then placed either on an EPM facing a closed arm or into the center of a circular open-field maze and observed for 15 min. Behavior in the mazes was recorded for 15 min, and anxiety-like behavioral parameters were quantified afterward with ANYmaze software.

Constant optogenetic pIC inhibition during acquisition of auditory fear conditioning. Mice were subjected to a fear conditioning and recall paradigm by using a standard fear conditioning chamber (Ugo Basile), consisting of a behavior box with transparent walls and an electric grid floor. Freezing as a measure of fear behavior was scored with ANYmaze. Mice were considered to be frozen if no movement except breathing was detected for at least 2 s. On day 1, mice were habituated in the morning to four presentations of the later conditioned stimulus (CS^+; pips at 7.5 kHz, 250 ms and 2 Hz for 30 s) and a control tone (CS^-; white noise for 30 s) with random ISIs over a period of 15 min. Fear conditioning was performed on the afternoon of the same day by pairing five CS^+ presentations with foot shocks (0.4 mA, 1 s in duration as an unconditioned stimulus, US). The CS^- tone was presented after each CS^+–US association but was never reinforced, and the ISIs were of random duration, lasting between 50 and 140 s. During the entire fear conditioning phase (21 min), the pIC was inhibited (593 nm, 10 mW). On day 2,

fear recall was tested without pIC inhibition in the same context, via four CS^+ and four CS^- presentations with random ISIs (total experiment duration of 16 min).

Fiber photometry. *Fiber photometry during the elevated plus maze and elevated zero maze.* Mice were handled for at least 3 d and habituated to being tethered. On the day of the experiments, mice were tethered to optic patch cords (low-autofluorescence fibers, 0.48 NA; Doric Lenses) and connected to the fiber photometry setup via a rotary joint. After 3 min of acclimatization in their home cage, the mice were tested for 15 min on the EZM or EPM. Videos of the mice were acquired and were analyzed with ANYmaze and synchronized to photometry data with TTL (transistor-to-transistor logic) pulses.

Photometric signal acquisition and analysis. As described previously[16], we used GCaMP6s bulk fluorescence measurements through a single optical fiber (custom made, 200 µm in diameter, 0.48 NA) for both delivery of excitation light and collection of emitted fluorescence. For both excitation and emission measurements, we used a commercial fiber photometry system (one site, two color; Doric Lenses) with two excitation wavelengths, 405 nm (isosbestic point of the GCaMP signal; calcium independent) and 465 nm (GCaMP signal; calcium dependent). The two excitation LEDs were fiber coupled into patch cords. Excitation light intensity was measured at the end of these patch cords as ~20 µW for 465-nm light and ~7 µW for 405-nm light. The patch cords were coupled to implanted optic fibers before the behavior experiments. Implanted fibers were tested before implantation to transmit ≥80% of the incoming light.

Photometry signals were recorded at 12 kHz, demodulated and downsampled to 30 Hz for analysis. All photometry data were analyzed with custom-written MATLAB programs. To remove the strongest photobleaching artifacts, we excluded the first 30 s of each recording. To remove bleaching and motion artifacts, a least-squares linear fit was applied to the 405-nm control signal and aligned to the 465-nm signal, by using a procedure developed by Lerner et al. (https://github.com/tallalerner/Photometry-Analysis-Shared/blob/master/Dropbox/MATLAB/Shared%20photometry%20code/controlFit.m).

The change in fluorescence ($\Delta F/F$) was calculated as $\Delta F/F$ = (465-nm signal − fitted 405-nm signal)/fitted 405-nm signal.

For the correlation analysis in Fig. 2d, we built z scores by using the mean and s.d. calculated across the entire behavioral session to make comparisons between mice feasible. We then computed the AUCs of the z scores for time spent in either the open or closed arms and plotted them on the y axes.

Approach and retreat analysis. For the EPM, we defined an approach as movement from the corner of a closed arm toward an open arm. Conversely, a retreat was defined as movement from an open arm toward the corner of a closed arm (see Fig. 5a,b for schematics). Additionally, we also considered movement from the corner of a closed arm toward the center as an approach, even when the mouse did not enter an open arm. To define approach and avoidance on the EZM, the position of the mouse was transformed from Cartesian coordinates to angular displacement ($-\sin(2\theta)$). The position of the mouse (θ) was computed by taking the dot product of two vectors, namely, a vector extending from the center of the EZM to the starting point of the mouse, which was defined as a closed arm–open arm transition (0_x in Supplementary Fig. 5b), and a vector extending from the center of the EZM to the position of the mouse on the EZM. The position (θ) was then converted to $-\sin(2\theta)$, such that the mouse's angular displacement was zero when the mouse was at the border between an open and closed arm, negative when the mouse was in an open arm and positive when the mouse was in a closed arm. The angular displacement was at its maximum of 1 when the mouse was at the extreme of a closed arm and −1 when the mouse was most exposed on an open arm.

To calculate the change in $\Delta F/F$ as plotted in Fig. 2h, we first annotated approaches and retreats manually. Approaches were only considered if they started at the end of a closed arm. Retreats were defined as any movements leaving an open arm. We first calculated the mean derivative of $\Delta F/F$ for each approach and retreat for each mouse. We then averaged all approaches and retreats made by a single animal to plot them as shown in Fig. 2h.

In vivo electrophysiology. Coordinates for optrode implantations in the pIC were the same as for the stereotactic surgeries for optogenetic experiments. Electrodes extended up to 500 µm below the optic fiber. Viruses encoding either ChR2 (AAV2/5-*Camk2a*-hChR2(H134R)-eYFP) or NpHR (AAV2/5-*Camk2a*-eNpHR3.0-eYFP) were unilaterally injected into the pIC, and a custom-built optrode was implanted above the injection site. Optrodes consisted of a 16-wire electrode array (Nichrome wires, 13 µm in diameter, impedance of 30–100 KΩ; Kanthal) combined with an optic fiber (200 µm in diameter, 0.22 NA), and the electrode array was attached to an 18-pin connector (Omnetics). The 16 electrode wires were cut at an angle such that all wires extended to slightly different depths below the optic fiber. Additionally, optrodes were referenced/grounded via a silver wire (127 µm in diameter; A-M Systems) placed above the cerebellum. For recordings, mice were tethered to an optical patch cord, connected to a laser (473 or 593 nm) and to an electrophysiological head stage, placed in a behavioral arena with transparent plexiglass walls and recorded with a webcam at 30 f.p.s. Neuronal activity data were band-pass filtered between 0.5 and 7.5 kHz and digitized at 30 kHz with a 16-channel amplifier system (RHD2000, Intan Technologies). For short-term inhibition experiments, NpHR-expressing mice received ten 5-s constant illuminations of yellow (593-nm) light (10–12 mW) with an ISI of 15 s. For long-term inhibition, after a baseline was recorded for 5 min, NpHR-expressing mice received constant yellow (593-nm) light for 20 min, followed by a 5-min post-light recording. For analysis of LFP activity during stimulation-evoked behaviors in ChR2-expressing mice, laser stimulation (473 nm) was manually triggered by the experimenter in a random fashion over 20 min (1 s, 20 Hz, 5-ms pulse width, 3–5 mW, minimum ISI of 4 s). The experiment was terminated prematurely if we observed severe seizure-like activity. Stimulation-induced behaviors were manually scored and classified from video recordings. For single-unit analysis, data were high-pass filtered at 250 Hz and referenced by subtracting the wide-band signal for one channel from all other channels. Spike detection and sorting were performed with Offline Sorter 4.4.0 software (Plexon) and data were analyzed with Neuroexplorer 5.112 (Nex Technologies). Waveforms were manually defined with guidance from three-dimensional visualizations of the principal components, timing and voltage features of the waveforms. A single unit was defined as a cluster of waveforms constituting a discrete, isolated cluster in the feature space that did not contain spikes with a refractory period of less than 1 ms on the basis of autocorrelation analyses. Unit isolation was verified with auto- and cross-correlation histograms. To analyze the effects of long-term inhibition, peristimulus time histograms (PSTHs) were computed for each neuron with 10-s bins and a z score was calculated by comparing the firing rate after light onset with the firing rate during the 5 min before light onset. The average firing rates of all neurons were compared between the light-on and light-off phases in 5-min bins. For short-term inhibition, PSTHs were computed for each neuron with 500-ms bins and a z score was calculated by comparing the firing rate after each onset of light with the firing rate during the 5 s before each onset. Furthermore, the average firing rates of all neurons were compared between the light-on and light-off phases in 5-s bins. LFP data were analyzed with MATLAB. Wide-band data from each mouse were referenced by subtracting the signal of one channel from all other channels. Light-stimulation-induced artifacts were removed from the signal by excluding the 1-s periods during which light stimulation occurred. LFP activity was quantified for each observed behavior by extracting 2–3 LFP traces of 5 s in duration for each mouse and behavior and averaging the absolute values. Seizure activity was classified as LFP activity of large amplitude (absolute amplitude of >0.2 mV) that lasted at least 5 s.

Heart and respiratory measurements in anesthetized mice. Anesthesia (isoflurane) was initiated at 5% and maintained at 1–1.5% during experiments. Mice were placed in a stereotaxic frame, and heart rate and breathing rate were recorded at 15 Hz with a MouseOx Plus oximeter (s-collar clip, shaved neck; Starr Life Sciences) while delivering bilateral optogenetic manipulations. For ChR2-expressing mice, this consisted of a 30-s baseline followed by a 20-s stimulation epoch of 5, 10 or 20 Hz at 473 nm, 3–5 mW with a 5-ms pulse width. Different frequencies were tested within the same mouse after a 2-min break. For NpHR-expressing mice, we recorded a 30-s baseline followed by a constant delivery for 2 min of 593-nm light (12 mW). z scores were calculated from the mean and s.d. for the first 20 s of the baseline recording.

Awake head-fixed two-photon imaging. Two-photon imaging was carried out on a rotating resonant scanning two-photon microscope (B-scope, Thorlabs) with a 16× water-immersion objective (Nikon, N16XLWD-PF; 0.8 NA) set to an angle perpendicular to the IC surface (75–85°). This provided a 415 × 415 µm² field of view that was scanned at 7.695 Hz with a resolution of 512 × 512 pixels. All imaged fields of view were imaged at a depth of 150–250 µm below the brain surface, by using a Mai Tai DeepSee laser (Newport Corp.) set to 940 nm and a power of 12–30 mW at the front aperture of the objective.

Experimental timelines. All mice were handled by the experimenter for a period of at least 3 d before any surgical procedures. We used a combination of C57BL/6NRj mice (12–18 weeks old) in which GCaMP6s was expressed from AAV9-hSyn-GCaMP6s viral vector injected by stereotaxic surgery 2 weeks before craniotomy and transgenic mice (tetO-GCaMP6s × *Camk2a*-tTA) expressing GCaMP6s in excitatory neurons ($CaMKII\alpha^+$). After craniotomy and head post implantation as described above, mice were allowed to rest for a period of 7–10 d. After this, they were handled by the experimenter and habituated to the two-photon imaging setup and head fixation for a period of 5 d. For fixation to the head holder on the setup, mice were briefly anesthetized with 5% isoflurane. After any head fixation, the mice were always allowed at least 30 min to recover and habituate. Mice were allocated to ExpA, ExpB or ExpC.

ExpA. Mice in this experiment underwent an additional 3 d of habituation to tail shock electrodes and training to lick from a spout positioned close to their mouth. On day 4, the pIC was imaged while mice were presented with a randomized set of quinine and sucrose solutions and painful tail shocks. To be consistent with previous studies that used 30–50 mM sucrose solutions[13,19] and at the same time not miss possible responsiveness, we chose a concentration of 60 mM (or 20%) sucrose in water. Quinine concentration was the same as during behavioral experiments (0.5 mM).

ExpB (fear conditioning). Mice in this experiment underwent an additional 3 d of habituation to tail shock electrodes. On day 4, activity in the pIC was imaged while mice underwent a fear conditioning protocol. This consisted of an initial habituation phase in which mice were presented with a 30-s CS^+ tone (7.5-kHz pips of 250 ms in length at 2 Hz) 5 times. Five to 10 min later, a training phase followed in which mice were presented with the CS^+ tone 6 times, with each tone coinciding with a 0.5-s tail shock at the end of the stimulus presentation. Sixty minutes after the conclusion of the training phase, a recall/extinction phase followed either with or without tones. For extinction with a tone, the CS^+ tone was presented 20 times, but it was not paired with any tail shocks. No imaging was performed between the fear conditioning phases.

ExpC. Mice in this experiment underwent an additional 3 d of training to lick from a spout positioned close to their mouth and habituation to intraperitoneal saline injections. On day 4, mice were acutely water deprived for 16–20 h and activity in the pIC was imaged while they underwent the following protocol. After a brief baseline recording, mice were presented with drops of sucrose solution for 10 min and were allowed to drink to satiation, followed by another baseline recording 25 min later. Fifteen minutes after the thirst/sucrose satiation experiment, we started the malaise experiment. After 5 min of baseline recording, mice were carefully injected intraperitoneally with a LiCl solution (0.15 M LiCl in 0.9% saline, 2% of body weight (for example, 600 µl for a 30-g mouse)) or only saline. They were then imaged for a period of 30 min to visualize pIC activity upon the development of malaise.

Taste deliveries. Tastants were presented to mice through a round-ended reusable feeding needle (Fine Science Tools) for 1.5 s at a time, controlled by using a solenoid pinch valve (225PNC2-21, NResearch). Delivered tastants were as follows: 20% sucrose solution or 0.5 M quinine solution (Sigma-Aldrich) and drinking water (from our animal facility).

Shocking. Painful stimuli were delivered via a weak electrical shock (0.5 s, 0.4 mA) to the tail generated by a Precision animal shocker (Coulbourn Instruments). Two 0.5-mm silver-coated copper wire coils (Conrad), custom made for each mouse, were soldered to a light-weight isolated cable connected to the shocker and positioned approximately 1 cm apart at the center of the tail.

Pupil and orofacial videography. We acquired data by using two USB3.0 monochrome cameras (BFS-U3-13Y3M-C, Point Grey Research), one positioned in front of the mouse and the other positioned perpendicularly to the side of its head, contralaterally to the cranial window. Illumination was provided by three 875-nm infrared LED arrays (Kemo Electronic, M120). Each camera was fitted with a near-infrared short-pass filter with a 900-nm cutoff (FES0900, Thorlabs) to filter out two-photon excitation light. To reduce baseline pupil size, mice were constantly and indirectly weakly illuminated from behind by a 470-nm fiber-coupled LED (M470F3, Thorlabs).

Pupil diameter and orofacial movement analysis. We measured pupil diameter and orofacial movements as previously described[13,14].

Freezing detection. All periods of immobility lasting at least 2 s, where *z*-scored pupil area was at least 1 s.d. above the *z*-scored orofacial movement, were classified as freezing episodes.

Neuronal response analyses. Two-photon acquired neural data were analyzed with the two-photon analysis toolbox Suite2p (https://github.com/cortex-lab/Suite2P; see also ref. [33]), which was used to perform motion correction, neuron detection, cell segmentation and fluorescence measurement over time for each cell. For each field of view, the cell segmentation was manually reviewed and corrected when necessary. Responding neurons were detected as described previously[13]. Specifically, for data from ExpA, $\Delta F/F$ was calculated from the formula $(F - F_0)/F_0$, where F was momentary cell fluorescence and F_0 was the mean baseline fluorescence during the 10 s before stimulus presentation. Next, s.d.(0), the standard deviation of the prestimulus baseline, was determined. Neurons were considered as responsive when $\Delta F/F$ exceeded 3.5 s.d.(0) above F_0 during a stimulus presentation. Responses for different phases during anxiety, malaise and thirst were determined as neurons whose mean *z*-scored fluorescence was significantly higher or lower in a relevant state than the baseline mean florescence (after extinction, before LiCl injection and after drinking). The significance of the difference was determined by non-parametric Mann–Whitney *U* test with the significance threshold placed at $P < 0.001$ by using the appropriate one-sided alternative hypothesis.

Assessment of overlapping representations. All cells activated by quinine, sucrose and water (6.8%, 48/630 cells) were assumed to encode for the orofacial sensory experience of drinking and were excluded from the quinine- and sucrose-responding cell populations in Fig. 4.

Histology and validation of injection sites and optic fiber placements. Mice were anesthetized with ketamine and xylazine and were perfused intracardially with 1× PBS followed by 4% paraformaldehyde (PFA) in PBS. Brains were postfixed for an additional 1–2 d in 4% PFA at 4 °C. Then, brains were sliced with a vibratome into 70-µm coronal sections. We collected all sections sequentially in single wells of 96-well plates. Every second coronal section was mounted and acquired with a modified epifluorescence microscope equipped with a controllable stage (Ludl) orchestrated by µManager 2.0 beta, and sections were stitched with the Grid/StitchCollection plugin in Fiji (ImageJ). We determined the placement of the optic fibers as well as the extent of viral infection by referencing a mouse brain atlas[34]. Only mice with correct fiber placement and viral spread confined to the pIC were included in the study.

Quantification of anatomical tracings. *Rabies analysis.* Images were preprocessed with the trainable Weka segmentation plugin for Fiji. Segmented GFP^+ cells were quantified with a custom-written Fiji macro. We created a collection of ROIs for each coronal section with guidance from a mouse reference atlas. Rabies-positive neurons located between +2.8 mm and −6.2 mm with respect to the bregma (anterior–posterior) were counted and anatomically annotated. Data are reported as cell counts normalized to total cell counts (percent of total input). TVA expression can be leaky[17,35], so starter cells might be confused with input cells. In control experiments (injection of AAV-FLEX-TVAmCherry and AAV-FLEX-RG followed by RABVΔG-GFP(EnvA) into C57BL6/NRj mice and injection of AAV-FLEX-TVAmCherry followed by RABVΔG-GFP(EnvA) into *Camk2a*-Cre mice), we observed a few dozen to a hundred GFP^+ cells around the injection site that were confined to the pIC (−0.11 to −1.22 mm). No long-range inputs were observed. As we were only interested in long-range connectivity and not in local microcircuitry of the posterior insula, we excluded all input cells of the pIC and claustrum between −0.11 and −1.22 mm from analysis.

CTB double-labeling analysis. Images of coronal sections were acquired with an epifluorescence slide scanner, as described above. CTB-positive cells were segmented with the trainable Weka plugin as described for the rabies analysis, with separate classifiers for the 488-nm and 555-nm signal. We determined cell counts within the rostro-caudal extent of the IC (AP: +2.45 mm to −1.22 mm) by using the same custom-written Fiji script as for the rabies analysis. Double-labeled cells were manually quantified with the Fiji cell counter plugin.

Axonal AAV tracing analysis. Axonal AAV tracings were acquired on a Leica SP5 confocal microscope. For each brain, we determined the densest efferents outside the IC (usually in the VPMpc or fundus of striatum (IPAC)) and adjusted the acquisition settings to obtain a nearly saturated signal for this region. We acquired a single optical *z* section of 10 µm in thickness and positioned it in the middle of the coronal slice (*z* axis). Every second brain section from +2.8 to −6.2 mm with respect to the bregma was acquired. Images were preprocessed with Hessian ridge detection as described elsewhere[36]. Briefly, this results in binarized images of $eYFP^+$ axons while eliminating background fluorescence. These binarized images were quantified with a custom-written Fiji macro. Similarly to the rabies quantification, we used our custom ROI atlas synthesized from a mouse brain reference atlas and manually adjusted the ROIs for each coronal section. We measured the area of the ROIs and determined the total pixel count per ROI. Data are reported as the 'percent of total output', which is the pixel count of an ROI normalized to the sum of all pixels identified as belonging to axons. Additionally, we calculated the 'percent innervation density', which is the proportion of pixels covering the total area of an ROI, averaged across all sections on which this ROI is present. We excluded passing fiber bundles (striatum, cerebral peduncles, anterior commissure, internal and external capsules, pyramidal tract) from the analysis. The spread of infected cell bodies in the pIC was determined and excluded from analysis of the pIC or claustrum.

Immunohistochemistry. Mice were stereotaxically injected with 150 nl of AAV2/5-*Camk2a*-eYFP in the pIC as described above. After 2 months, mice were perfused with ~50 ml of room-temperature 1× PBS with heparin (10,000 U ml^{-1}), followed by 50 ml of 4% PFA in 1× PBS. Brains were removed and postfixed for 48 h at 4 °C. Coronal sections were cut with a vibratome at 50-µm thickness. At every step, we used solutions that did not contain a permeabilizing agent[37]. Sections were blocked for 2 h in blocking buffer (5% normal goat serum and 5% BSA in PBS) and incubated with mouse anti-Gad67 antibody (1:200 dilution; clone 1G10.2, Merck Millipore) at room temperature for 48 h. After three washes in 1× PBS, sections were incubated in Alexa Fluor 647-conjugated anti-mouse secondary antibody (1:200 dilution; Molecular Probes). Sections were mounted and imaged on an Axioobserver Z1 inverted microscope (Zeiss) with a 20× objective (PLAN-neofluar 20× objective, 0.5 NA; Zeiss) equipped with a CSU-X1 spinning disc confocal unit (Yokogawa Electric Corporation) controlled by VisiView software (Visitron Systems) and a CoolSnapHQ2 CCD camera (Photometrics).

Statistical analysis. Analyses were performed with GraphPad Prism (GraphPad Software, version 8), MATLAB (Mathworks) or Python. Group comparisons were made by either one-way or two-way ANOVA followed by Bonferroni post hoc tests if a significant main effect or interaction was observed ($P < 0.05$). Single-variable

comparisons were made with two-tailed unpaired or paired *t*-tests. Correlation analyses were made using linear regression. Responding cells in two-photon experiments were detected by using a *z*-score threshold (3.5 s.d.; ExpA) or a Mann–Whitney *U* test for longer-lasting states (ExpB and ExpC) with the significance threshold set at $P < 0.001$. Numbers of mice for all experiments are reported in the figures and their legends. Data distribution was assumed to be normal, but this was not formally tested. Only for two-photon and related pupil analyses did we not assume normality and tested the distribution by Mann–Whitney *U* test.

Sample sizes. No statistical methods were used to predetermine sample size for single experiments, but the sample sizes were similar to or greater than those reported in previous studies related to our experimental assays[5,11,25,27,30,44].

Data exclusions. We excluded mice post hoc when optic fiber placement or viral expression patterns were not appropriate (outside the target region, viral expression too weak) or when implants were lost during tests lasting for multiple days. Mice that fell off the EPM or EZM were excluded. Mice in fiber photometry experiments were excluded when the signal-to-noise ratio was in the range of the signal for eYFP controls. For whole-brain anatomical analyses, we excluded all input cells of the IC and claustrum located between −0.11 and −1.22 mm. The tracing analysis further included only coronal sections from +2.45 to −6.11 mm with respect to the bregma to allow reliable mounting and cutting. Measurements were excluded from the physiological recordings of anesthetized mice when anesthesia was not stable throughout the recording or when sensors lost signal.

Replication, randomization and blinding. All experiments reported here were reliably reproduced in independent experimental groups for behaviors or several mice for calcium imaging and viral tracings. For all behavioral tests and physiological recordings, littermates were randomly assigned to the different groups. Data collection and analysis were not performed with blinding to the conditions of the experiment if not otherwise stated in the methods section. Computer-based analyses ensured unbiased data collection and analysis in most cases. Post hoc video scoring by human raters was performed with blinding whenever possible.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

All custom-written analysis code is available from the corresponding author upon reasonable request.

References

50. Wekselblatt, J. B., Flister, E. D., Piscopo, D. M. & Niell, C. M. Large-scale imaging of cortical dynamics during sensory perception and behavior. *J. Neurophysiol.* **115**, 2852–2866 (2016).
51. Tye, K. M. et al. Dopamine neurons modulate neural encoding and expression of depression-related behaviour. *Nature* **493**, 537–541 (2013).
52. Labouèbe, G., Boutrel, B., Tarussio, D. & Thorens, B. Glucose-responsive neurons of the paraventricular thalamus control sucrose-seeking behavior. *Nat. Neurosci.* **19**, 999–1002 (2016).
53. Felix-Ortiz, A. C. et al. BLA to vHPC inputs modulate anxiety-related behaviors. *Neuron* **79**, 658–664 (2013).
54. Gardner, M. P. H. & Fontanini, A. Encoding and tracking of outcome-specific expectancy in the gustatory cortex of alert rats. *J. Neurosci.* **34**, 13000–13017 (2014).
55. Pachitariu, M. et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. Preprint at *bioRXiv* https://www.biorxiv.org/content/10.1101/061507v2 (2017).
56. Franklin, K. B. J. & Paxinos, G. *Paxinos and Franklin's The Mouse Brain in Stereotaxic Coordinates* 4th edn (Academic Press, 2012).
57. Do, J. P. et al. Cell type-specific long-range connections of basal forebrain circuit. *eLife* **5**, 1–18 (2016).
58. Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A. & Uchida, N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* **74**, 858–873 (2012).
59. Grider, M. H., Chen, Q. & Shine, H. D. Semi-automated quantification of axonal densities in labeled CNS tissue. *J. Neurosci. Methods* **155**, 172–179 (2006).
60. Taniguchi, H. et al. A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**, 995–1013 (2011).

nature neuroscience

SUPPLEMENTARY INFORMATION

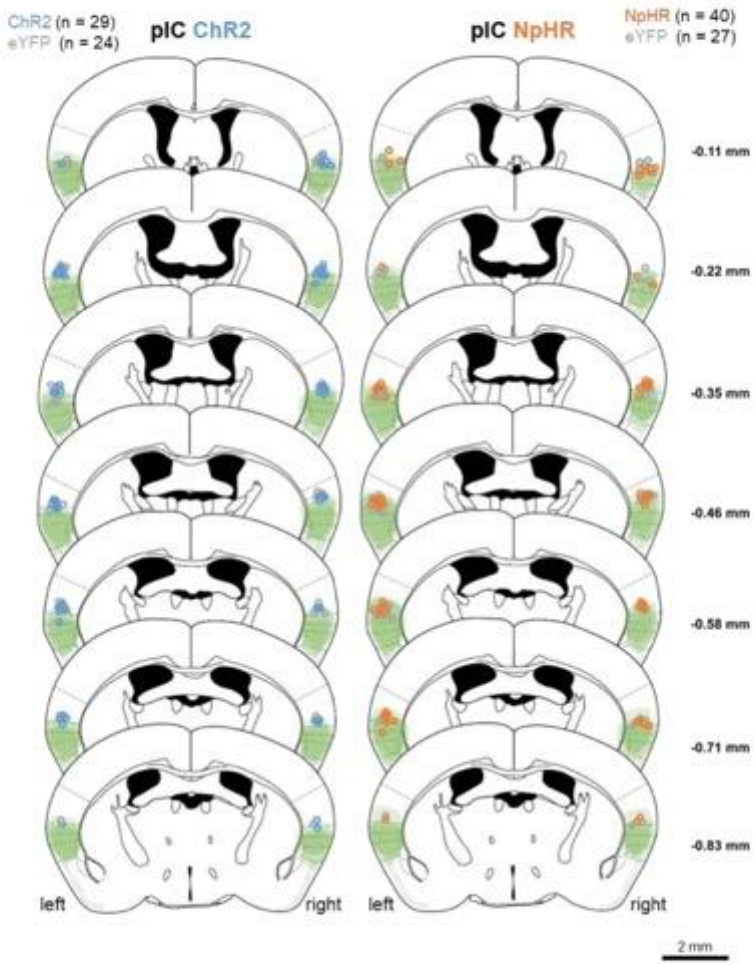
https://doi.org/10.1038/s41593-019-0469-1

In the format provided by the authors and unedited.

Aversive state processing in the posterior insular cortex

Daniel A. Gehrlach[1,2], Nejc Dolensek[1,3], Alexandra S. Klein[1,2], Ritu Roy Chowdhury[1,2], Arthur Matthys[1], Michaela Junghänel[1,5], Thomas N. Gaitanos[1], Alja Podgornik[1,2], Thomas D. Black[1,2], Narasimha Reddy Vaka[1], Karl-Klaus Conzelmann[4] and Nadine Gogolla[1*]

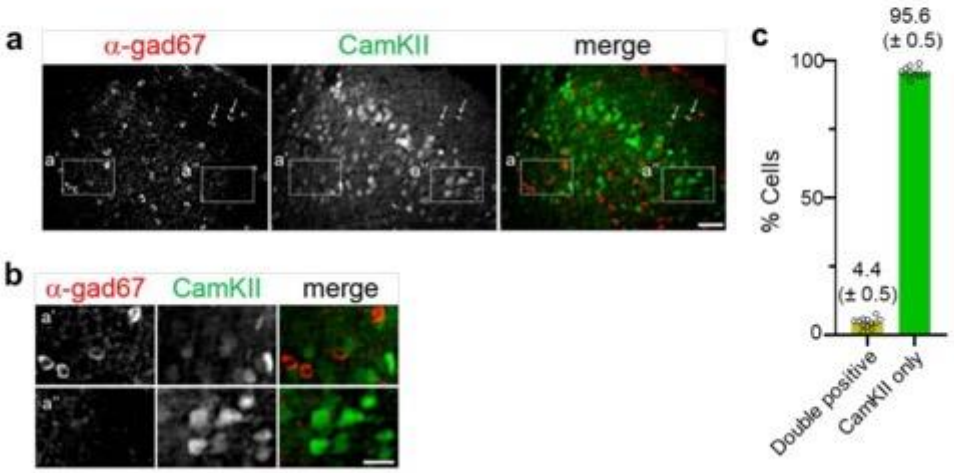
[1]Circuits for Emotion Research Group, Max Planck Institute of Neurobiology, Martinsried, Germany. [2]International Max Planck Research School for Molecular Life Sciences, Munich, Germany. [3]Graduate School of Systemic Neurosciences, Ludwig Maximilians University, Munich, Germany. [4]Max von Pettenkofer Institute and Gene Center, Medical Faculty, Ludwig Maximilians University, Munich, Germany. [5]Present address: Ausbildungsinstitut für Kinder- und Jugendlichenpsychotherapie an der Uniklinik Köln (AKiP), Cologne, Germany. *e-mail: ngogolla@neuro.mpg.de



Supplementary Figure 1

Histological verification of injection- and implantation sites for pIC-ChR2 and pIC-NpHR mice.

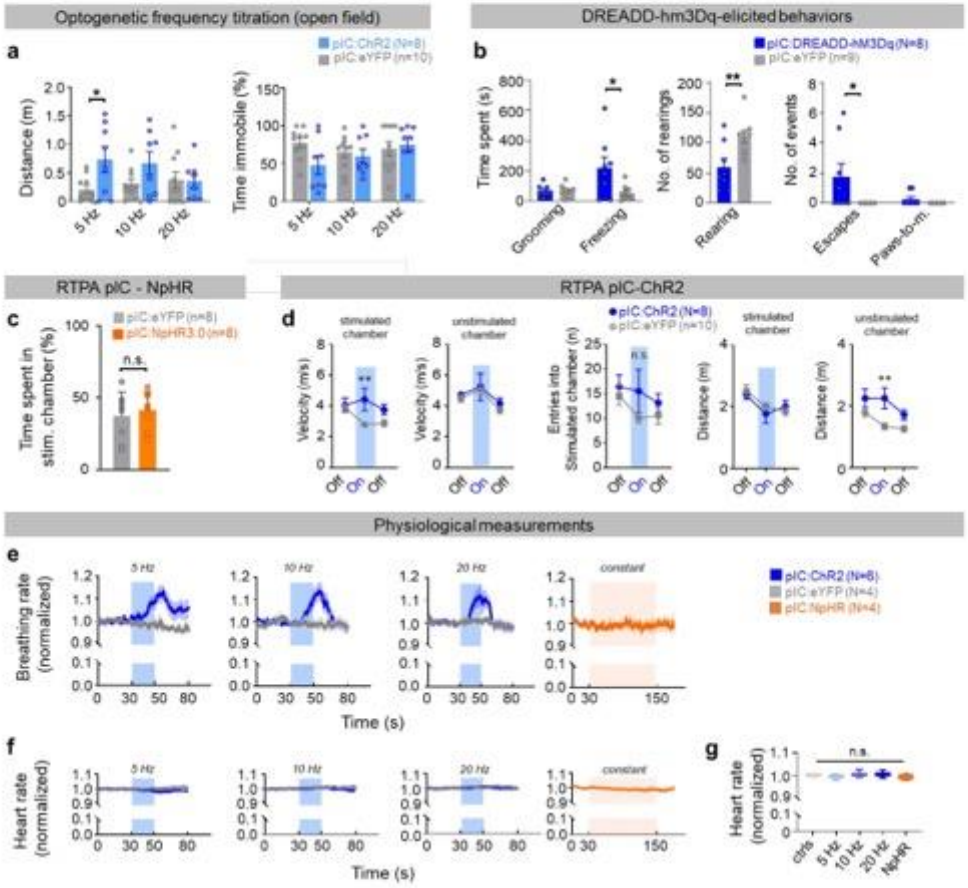
Related to Figures 1-3. Schematic visualization of ChR2 or NpHR expression (green-shaded regions). eYFP expression is not shown. Pooled data of three independent cohorts for each condition. Colored dots (blue: ChR2; orange: NpHR) represent the center of tip of the implanted optic fibers for ChR2 or NpHR expressing animals. Implantation sites for eYFP control animals are shown in grey.



Supplementary Figure 2

Histological verification of CaMKIIα-promoter-specificity.

Injection of 150 nl AAV2.5-CaMKIIα-eYFP into pIC followed by antibody staining against Gad67 (1:200 anti-Gad67, clone 1G10.2, Merck Millipore) to control for the specificity of the CamKIIα (1.2 kb) promotor used in this study. **a**, Representative image of the Gad67 staining. White arrows point at double positive neurons. Scale bars: 50 µm. **b**, Magnification of inserts a' and a'' of images in panel a. Scale bars: 25 µm bottom. **c**, Quantification of co-labelling. 4.4 % ± 0.5 s.e.m. of eYFP+ neurons stained for the inhibitory marker GAD67. Data was collected from 12 ROIs (n = 12 images) from four histological sections of two mice. In total 2175 cells were counted out of which 100 cells were double positive. Bar graphs indicate mean + s.e.m.

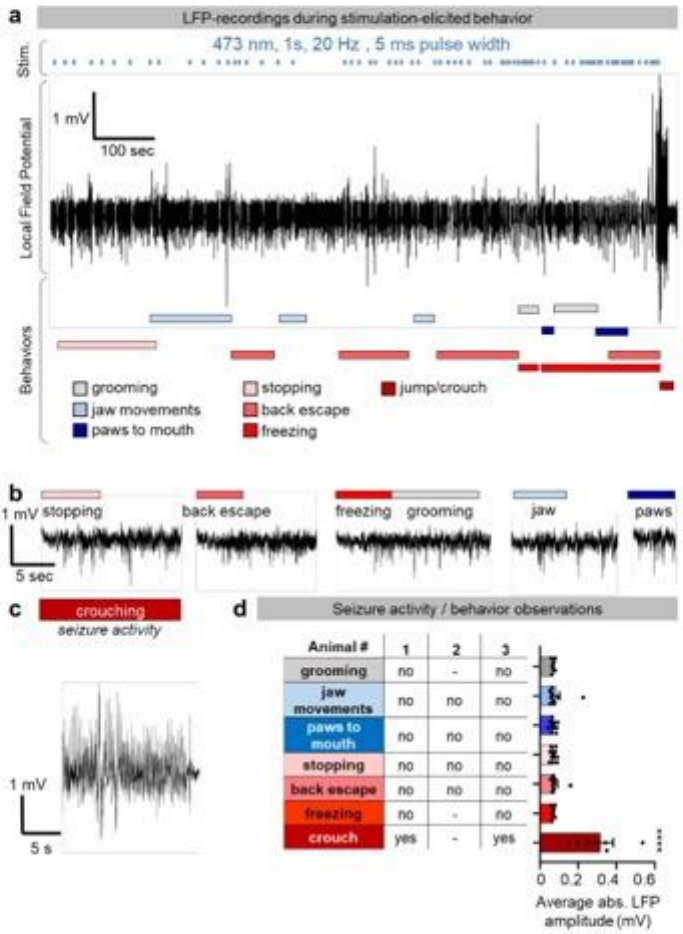


Supplementary Figure 3

pIC activation control experiments, RTPA and physiological measurements.

a, To titrate optogenetic stimulations we tested for frequency-dependent effects on locomotion. At 5 Hz, but not at 10 or 20 Hz, there was a significant increase in distance travelled (n = 8 ChR2 and n = 10 eYFP mice, two-way RM ANOVA, $F(1,16) = 5.283$, $p = 0.0353$, Bonferroni post-hoc analysis, eYFP vs ChR2 *p = 0.0411). We did not detect any overall change in the time spent immobile between the opsin and eYFP-expressing control group (two-way RM ANOVA, $F(1, 16) = 1.442$, $p = 0.2473$). **b**, To test whether stimulation-elicited behaviors as described in Figure 1 b-c also occur upon DREADD-mediated pIC activation, mice expressing the excitatory DREADD-receptor hM3Dq (n = 8) or eYFP (n = 9) in the pIC were injected with clozapine-N-oxide (CNO). Upon CNO injection, DREADD expressing mice displayed increased freezing (two-tailed unpaired *t* test, $t = 2.425$, $df = 15$, *p = 0.0284) and back escapes (unpaired two-tailed *t* test, $t = 2.165$, $df = 15$, *p = 0.0469), whereas the number of rearing events was reduced in comparison to control mice (two-tailed unpaired *t* test, $t = 2.959$, $df = 15$, **p = 0.0098). Paws-to moth movements were elicited in a subset of DREADD-expressing mice and never in controls, however these changes were not significant (two-tailed unpaired *t* test, $t = 1.627$, $df = 15$, $p = 0.1246$). Grooming remained unaffected (two-tailed unpaired *t* test, $t = 0.2308$, $df = 15$, $p =$

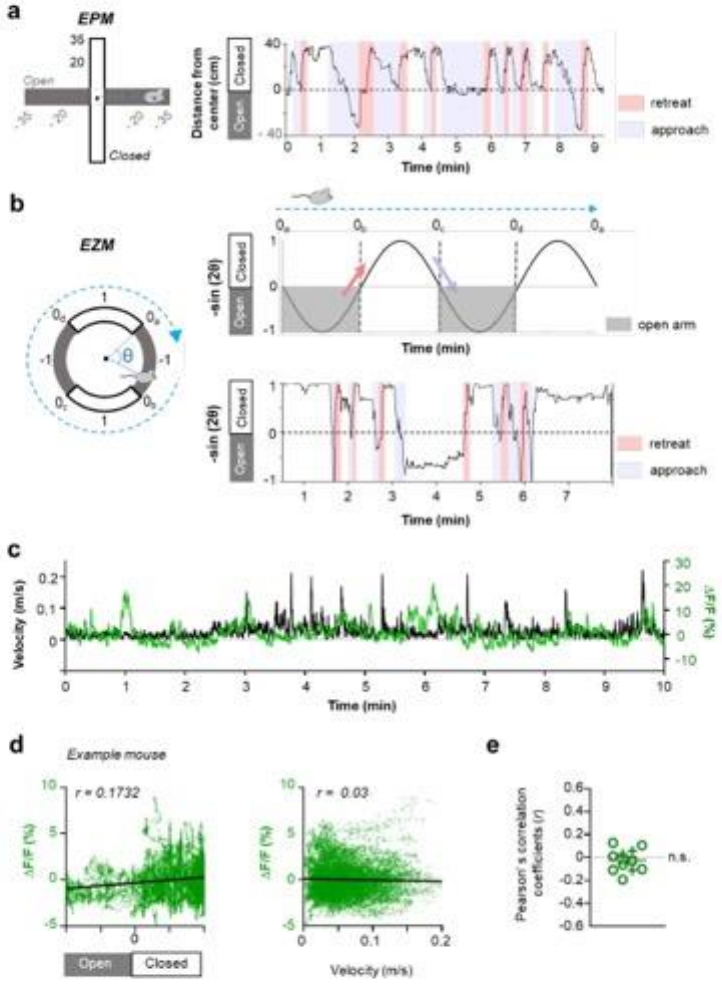
0.8206). **c**, Inhibition of the pIC did not affect real-time place preference. Time spent in the laser chamber on the day of laser stimulation was not significantly different between eYFP- and NpHR- expressing animals (n = 8 mice / group, two-tailed unpaired *t* test, t = 0.584, df = 14, p = 0.5685). **a-c**, bar graphs indicate mean + s.e.m. **d**, Analysis of additional locomotion parameters of the RTPA assay (n = 8 ChR2, n = 10 eYFP mice). Left: stimulation of pIC increased velocity in the stimulated chamber (two-way RM ANOVA, F (1, 16) = 5.225, p = 0.0362, Bonferroni post-hoc, **p=0.0099 eYFP vs. ChR2 on laser 'on' day), but not in the unstimulated chamber (two-way RM ANOVA, F (1, 16) = 0.8818, p = 0.3617). Middle: Stimulation does not affect the number of entries into the stimulated chamber (two-way RM ANOVA, F (1, 16) = 2.124, p = 0.1644). Right: stimulation of pIC did not affect the distance travelled in the stimulated chamber (two-way RM ANOVA, F (1, 16) = 0.2085, p = 0.6540), but increased the distance travelled in the unstimulated chamber (two-way RM ANOVA, F (1,16) = 9.772, p = 0.0065, Bonferroni post-hoc, **p=0.0056 eYFP vs. ChR2 on laser 'on' day). Lines are mean + s.e.m. **e-g,** Data from (n = 6 ChR2, n = 4 eYFP, n = 4 NpHR mice). **e**, Timeline of respiratory responses upon 5, 10, or 20 Hz stimulations or NpHR mediated inhibition of anesthetized mice as quantified in Fig. 1e Line graphs are mean with shaded areas s.e.m. **f**, Heart rate measurements upon ChR2-, or NpHR-mediated activity manipulations in the same animals as in panel e. Data are normalized to a 20-second baseline period for each individual. Line graphs are mean with shaded areas s.e.m. **g**, Quantitative assessment of the same data as in panel f. No significant difference between groups was detected (one-way ANOVA F (4, 33) = 2.872, p = 0.0381). Box-whisker plots display median, 25th to 75th percentiles, and min to max values.



Supplementary Figure 4

***In vivo* electrophysiological LFP recordings of the pIC during photostimulation-elicited aversive behaviors.**

a, Example LFP trace from one ChR2-expressing animal during unilateral photostimulation-induced behaviors. Laser activations (473 nm, 1 s, 20 Hz, 5 ms pulse width, minimum of 4 s ISI) are indicated above the trace (blue) and observed behaviors are shown in color code below. Note the high amplitude LFP signal towards the end of the trace. **b,** Example LFP traces from different photostimulation-induced behaviors. **c,** Only during crouching behavior, LFP recordings showed high amplitude (>0.2 mV absolute) signal with long durations (< 5 s) ("seizure-like" activity). **d,** Quantification of LFP recordings in three ChR2-expressing animals. LFP amplitudes were significantly increased during crouching but not during any other observed behaviors (data from 3 independent mice collectively exhibiting n = 6 grooming, n = 9 jaw movement, n = 9 paws-to-mouth, n = 9 stopping, n = 8 back escape, n = 6 freezing and n = 5 crouching episodes; one-way ANOVA, $F(6, 45) = 15.16$, $p<0.0001$, Bonferroni post-hoc analysis revealed that LFP amplitudes were significantly higher during crouching behavior compared to any other behavior, each comparison: ****$p < 0.0001$. Bar graphs indicate mean + s.e.m.

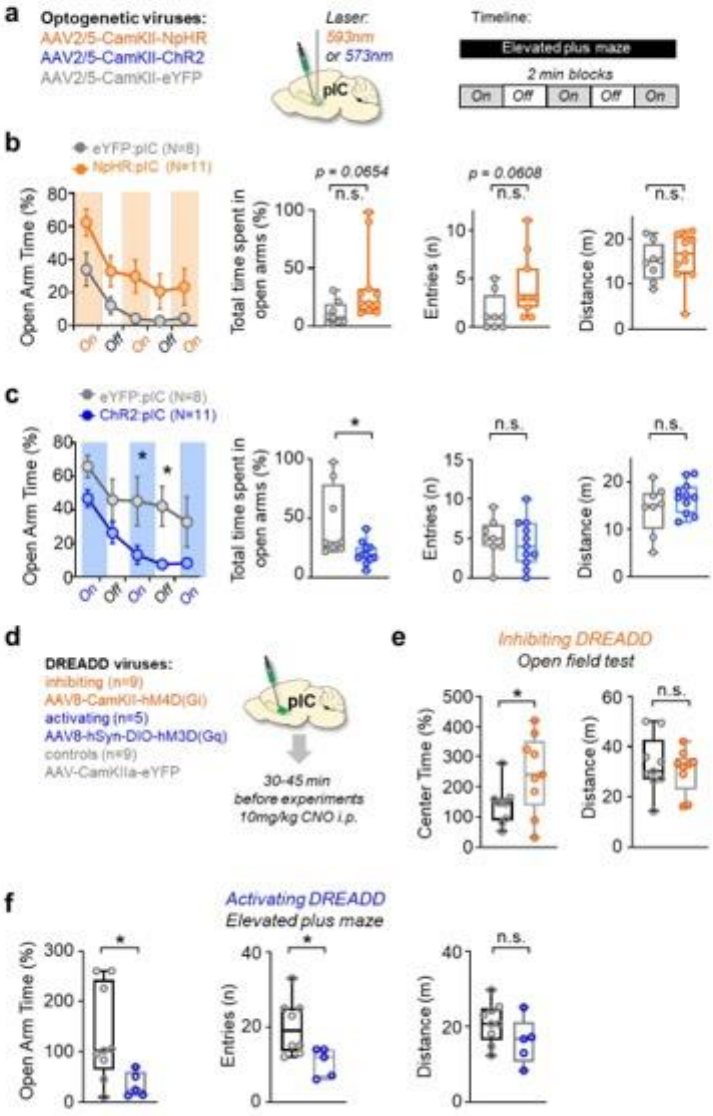


Supplementary Figure 5

Analysis of fiber photometry data by mobility in the elevated mazes.

a, Schematic illustration of approach / retreat behaviors in the elevated plus maze. Left: To segment behavior into approaches or retreats, we plotted the distance of an animal to the center point of the elevated-plus maze. On the open arms, the values would be negative, in the center zero, and in the closed arms positive. We defined approaches as movements from a closed arm towards an open arm. Conversely, a retreat was defined as movement from an open arm towards the closed arm. Right: example trace with annotated approaches (blue) and retreats (red). **b**, Left: In order to segment approaches and retreats on the elevated-zero maze (EZM), the position of the animal was transformed from Cartesian coordinates to angular displacement (-sin2θ). The position of the animal (θ) was computed from the dot product of the two vectors, namely, the vector between center of EZM to the starting point (define as point 0_a), and the vector between the center of EZM to the position of the animal on EZM. Right, top: The position(θ) was then converted

to -sin2θ, so that when the animal was at the borders between open and closed arms, it's angular displacement is zero, negative when in the open arms, and positive when in the closed arms. The angular displacement is at its maximum of 1 when the animal is at the extremes of the closed arm and (-1) when it is most exposed on the open arm. Right, bottom: Example trace of an animal exploring the EZM over time. Annotated approaches (blue) and retreats (red) are displayed. **c**, Example trace of the relationship between plC fluorescence and velocity in the elevated zero maze. **d**, Pearson's correlations of plC fluorescence as a function of position in the maze (left) and velocity (right) in one example animal. Left: Note that the activity is higher in closed arm and is positively correlated. Right: Velocity is very weakly correlated to plC activity. **e**, There is no consistent correlation of plC activity and velocity across animals (n = 11 mice, circles tested in EZM, pluses tested in EPM (same as in Fig. 2g), two-tailed one sample *t* test revealed no significant difference from zero, t = 0.8450, df = 10, p = 0.4179).

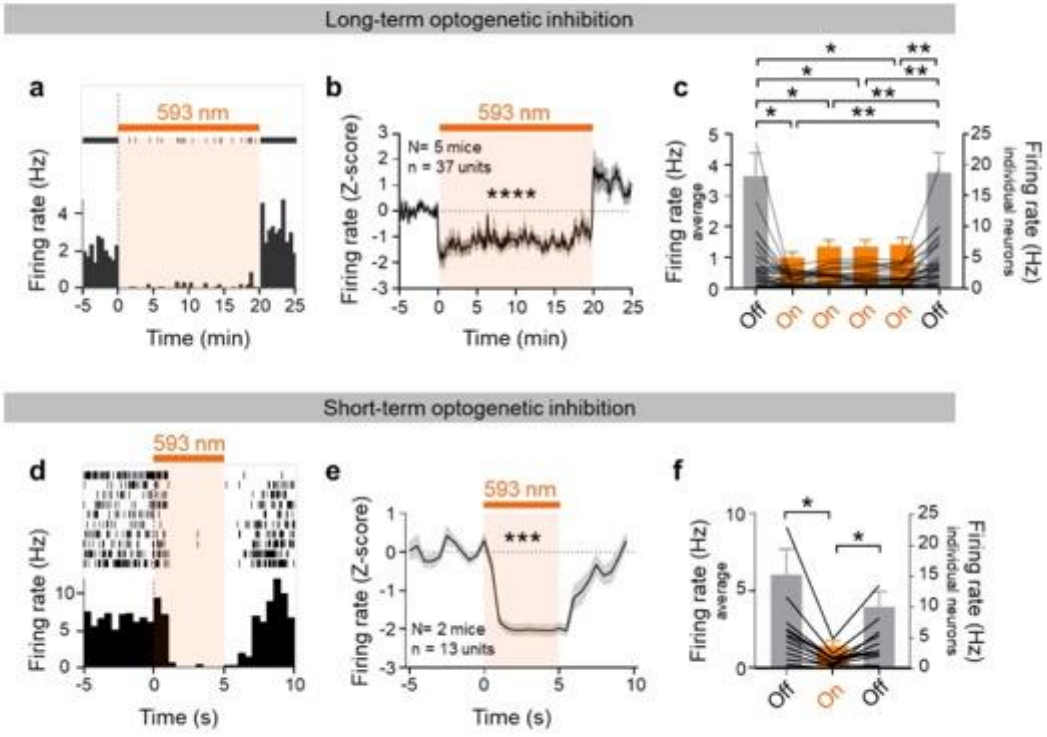


Supplemental Figure 6

Influence of optogenetic and chemogenetic plC inhibition or activation on innate fear behavior.

a, Experimental strategy and timeline for optogenetic plC activity manipulations using a fixed regimen of alternating 2-min laser 'on' and 'off' epochs. **b**, Data from n = 8 eYFP- and n = 11 NpHR-expressing mice. Left: Inhibition of plC activity did not significantly alter the time spent in the open arms, only a trend towards anxiolysis was observed (two-

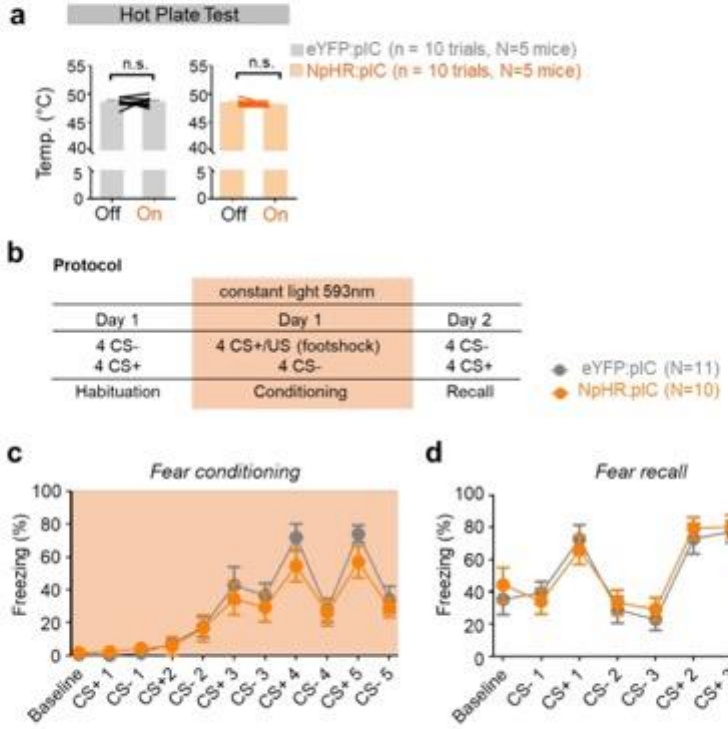
way RM ANOVA, group (opsin) effect, $F(1, 17) = 3.881$, $p = 0.0654$; time effect, $F(1.585, 26.95) = 24.73$, $p < 0.0001$, group x time interaction $F(4, 68) = 0.5624$, $p = 0.6907$). Right: Also the overall time spent in the open arms showed a trend to more exploration upon pIC inhibition (two-tailed unpaired *t* test: $t = 1.970$, $df = 17$, $p = 0.0654$). Neither the entries into the open arms (two-tailed unpaired *t* test, $t = 2.008$, $df = 17$, $p = 0.0608$), nor the total distances travelled within the maze (two-tailed unpaired *t* test, $t=0.4976$, $df=17$, $p = 0.6251$) differed between the groups. **c**, Data from n = 8 eYFP and n = 11 ChR2 -expressing mice. Left: Increasing pIC activity in alternating blocks of 2 min laser 'on' and 'off' epochs decreased the overall time spent in the open arms gradually and in a sustained manner (two-way RM ANOVA, group (opsin) effect, $F(1, 17) = 7.243$, $^*p = 0.0155$; time effect, $F(4, 68) = 12.74$, $p < 0.0001$, group x time interaction $F(4, 68) = 0.8126$, $p = 0.5215$; Bonferroni post-hoc tests revealed significant differences between groups in the 2^{nd} on and 2^{nd} off epochs, $^*p = 0.0450$ and $^*p = 0.0235$ respectively). Right: The overall time spent in the open arms throughout the 'on' and 'off' periods was significantly reduced (two-tailed unpaired *t* test, $t = 2.691$, $df = 17$, $^*p = 0.0155$). Neither the entries into the open arms (two-tailed unpaired *t* test, $t = 0.4832$, $df = 17$, $p = 0.6351$) nor the total distances travelled within the maze (two-tailed unpaired *t* test, $t = 1.254$, $df = 17$, $p = 0.2267$) differed between the groups. **d**, Viruses and strategy used to manipulate pIC activity using DREADDs. **e,** Constant silencing of the pIC throughout the exploration of the circular open field using DREADD-hM4D resulted in anxiolytic effects (n = 9 mice / group) as revealed by increased time spent in the center (two-tailed unpaired *t* test, $t = 2.165$, $df = 16$, $^*p = 0.0459$), but no difference in the total distance travelled (two-tailed unpaired *t* test, $t = 0.5524$, $df = 16$, $p = 0.5883$). **f**, Constant activation of the pIC throughout the exploration of the elevated plus maze using DREADD-hM3D resulted in anxiogenic effects (n = 9 eYFP mice, n = 5 DREADD mice) as revealed by a decrease in the total time spent (two-tailed unpaired *t* test, $t = 2.263$, $df = 12$, $^*p = 0.0430$) and fewer entries into the open arms (two-tailed unpaired *t* test, $t = 2.612$, $df = 11$, $^*p = 0.0242$), but no difference in the total distance travelled (two-tailed unpaired *t* test, $t = 1.565$, $df = 12$, $p = 0.1436$). Line graphs indicate mean + s.e.m. Box-whisker plots display median, 25^{th} to 75^{th} percentiles, and min to max values, as well as individual values.



Supplementary Figure 7

***In vivo* electrophysiological single-unit recordings in the pIC during long- and short-term optogenetic inhibition.**

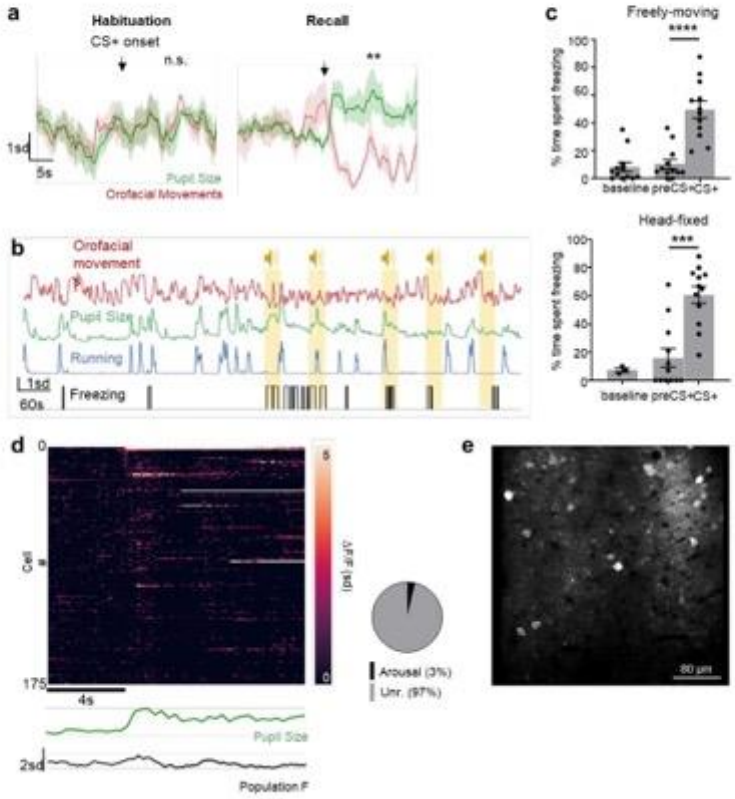
a - **c,** Long-term optogenetic inhibition in mice expressing NpHR. **a,** Reduced firing rate of a representative pIC neuron during 20 min inhibition (bin size = 30 s). **b,** Average z-scored firing rates from 37 single units recorded in 5 mice reveal significantly reduced activity during 20 min inhibition (bin size = 10 s, two-tailed one sample *t* test with theoretical mean = 0, t = 42.95, df = 119, ****p < 0.0001). **c,** Average and individual firing rates of 37 pIC neurons in 5 min bins are significantly reduced during 20 min inhibition (n = 37 neurons, one-way RM ANOVA F(1.814, 65.29) = 12.37, p < 0.0001, Bonferroni post-hoc analysis revealed significant differences pre- or post-inhibition: 1st Off vs. 1st On, *p = 0.0178; 1st Off vs. 2nd On *p = 0.0400; 1st Off vs. 3rd On *p = 0.0272; 1st Off vs. 4th On *p = 0.0348; 1st On vs. 2nd Off **p = 0.0018; 2nd On vs. 2nd Off **p = 0.0029; 3rd On vs. 2nd Off **p = 0.0023; 4th On vs. 2nd Off **p = 0.0017). **d-f,** Short-term optogenetic inhibition in mice expressing NpHR. **d,** Peri-stimulus time histogram (PSTH) reliably indicates reduction of firing rates of a representative pIC neuron during 10 repetitions of 5 s long inhibitions (bin size = 500 ms). **e,** Average z-scored firing rates from 13 single units recorded in 2 mice reveal significantly reduced activity during 5 s long inhibitions (bin size = 500 ms, two-tailed one sample *t* test with theoretical mean = 0, t = 5.967, df = 9, ***p = 0.0002). **f,** Average and individual firing rates of 13 pIC neurons in 5 s bins are significantly reduced during 5 s photoinhibitions (n = 13 neurons, one-way RM ANOVA F(1.295, 15.54) = 11.10, p = 0.0026, Bonferroni post-hoc: significant for pre vs. during inhibition, *p = 0.0115, and post versus during inhibition, *p = 0.0142). Line and bar graphs indicate mean + s.e.m.



Supplemental Figure 8

Constant pIC inhibition does not affect pain thresholds, nor does it affect acute fear expression and subsequent contextual or cued fear recall.

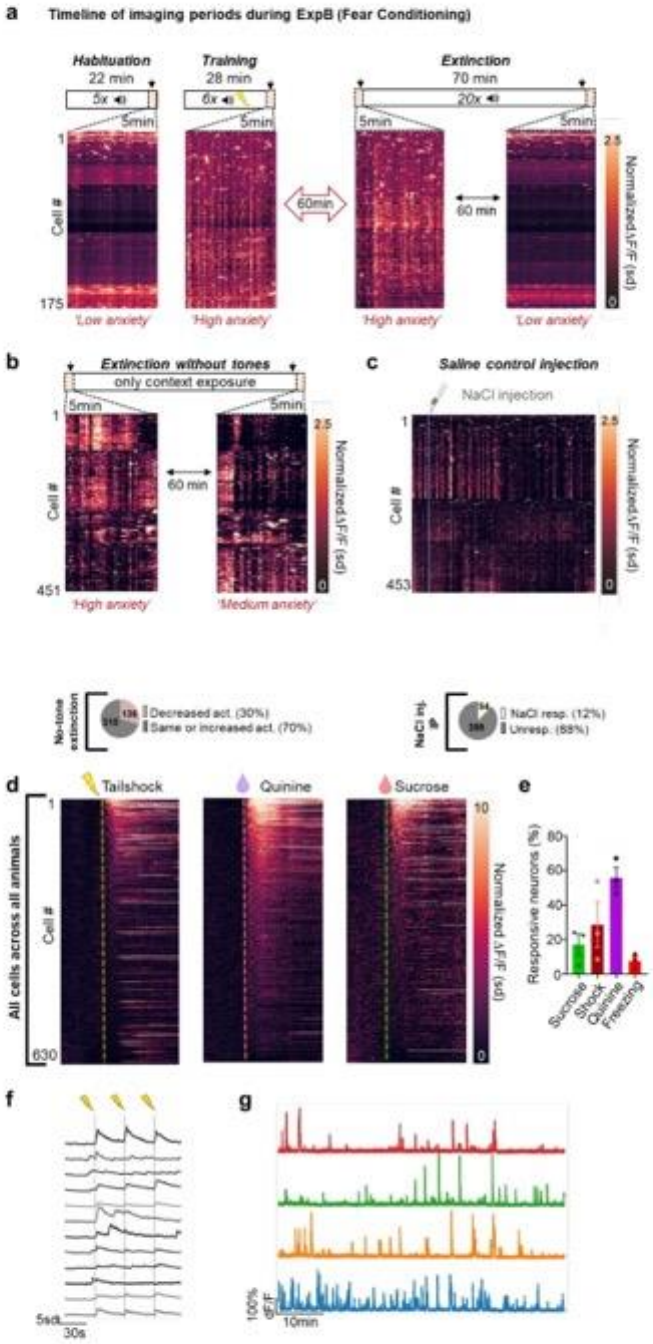
a, Ramping hot plate test (nociception control for effects in Fig. 3b-f). Temperature thresholds did not change for NpHR-expressing or eYFP-expressing mice (N = 5 mice / group tested twice each) between light 'off' and 'on' conditions (n = 10 trials in 5 eYFP mice: two-tailed paired *t* test, t = 0.02949, df = 9, p = 0.9771; n = 10 trials in 5 NpHR mice: two-tailed paired *t* test, t = 1.967, df = 9, p = 0.0807). Data is shown as means + s.e.m. **b**, Protocol for a separate auditory fear conditioning experiment conducted under permanent inhibition of the pIC during the conditioning phase. **c-d,** Data from n = 11 eYFP and n = 10 NpHR expressing mice. **c**, Constant pIC inhibition during auditory fear conditioning does not change freezing behavior (two-way RM ANOVA, group (opsin) effect, $F(1, 19) = 0.6106$, $p = 0.4442$, time effect, $F(4.944, 93.93) = 36.17$, $p < 0.0001$, opsin x time interaction, $F(10, 190) = 0.8369$, $p = 0.5937$. **d**, Neither contextual (during baseline) nor cued fear (during CS+ presentations) recall are altered subsequently to constant pIC inhibition during the conditioning (two-way RM ANOVA, group (opsin) effect, $F(1, 19) = 0.001852$, $p = 0.9661$, time effect, $F(5.462, 103.8) = 37.83$, $p < 0.0001$, opsin x time interaction, $F(8, 152) = 1.110$, $p = 0.3593$. Line graphs indicate mean + s.e.m.



Supplementary Figure 9

Detection of freezing in head-fixed mice.

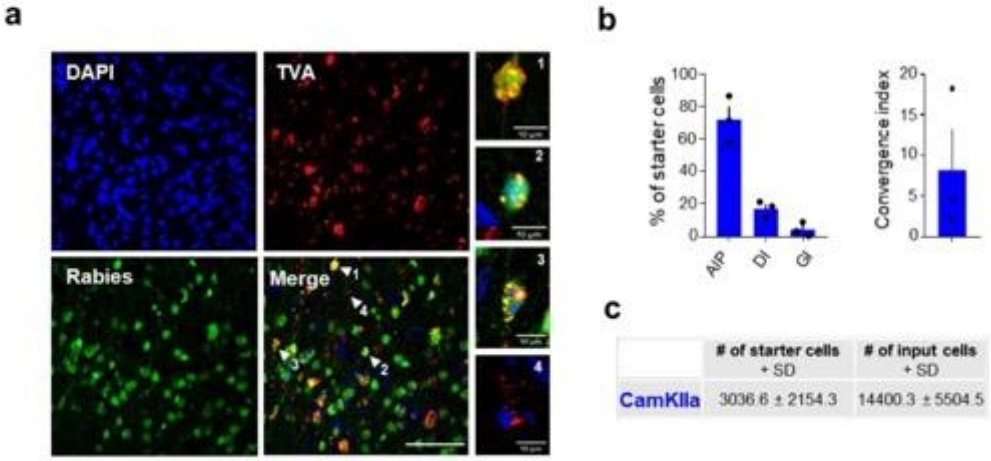
a, CS+ onset specifically during fear recall (right) but not during habituation (left) results in a strong increase in pupil size and intermittent stopping of orofacial movements (lines = mean z score; shaded areas = s.e.m.; two-tailed Mann-Whitney *U* test: pupil size versus orofacial movement score after CS+ onset (30 s). Habituation: $U = 12$, $p > 0.9999$; recall: $U = 0$, $^{**}p = 0.0079$; $n = 5$) which taken together define freezing in the described method. Lines are mean, shaded areas = s.e.m. **b**, Simultaneous tracking of running (blue), pupil size (green) and standardized orofacial movement score (red), and the resulting detected freezing episodes (black) for one representative mouse undergoing fear recall. Presentations of the conditioned auditory stimuli are shown in yellow. **c**, Freezing detected in head-fixed animals by our described novel technique, detects similar freezing levels under head-fixed conditions as in a conventional conditioning chamber using the commercial (ANY-maze, Stoelting) freezing detection module (4CS+ presentations in 3 head-fixed mice, n =12, two-tailed paired *t* test pre-CS+ versus CS+, $t = 5.874$, $df = 11$, $^{***}p = 0.0001$; $n = 12$ freely-moving C57BL6/N mice, two-tailed paired *t* test pre-CS+ versus CS+, $t = 9.483$, $df = 11$, $^{****}p < 0.0001$). **d**, Moment-by-moment analysis of neuronal activity in the pIC and pupil size. Non-freezing spontaneous arousal episode after 4 s (as indicated by pupil dilation; green trace) is not associated with pIC activity. Bottom: mean population fluorescence standard deviations (sd). All neurons whose fluorescence was at least 3.5 sd above baseline during increasing pupil size were considered as responding (N = 2 mice, n = 352 cells). **e**, Field of view in two-photon imaging of pIC. Bar graphs indicate mean + s.e.m.


a
Timeline of imaging periods during ExpB (Fear Conditioning)
Habituation
22 min
5x
Training
28 min
6x
Extinction
70 min
20x
5min
60min
60 min
Cell #
1
175
'Low anxiety'
'High anxiety'
'High anxiety'
'Low anxiety'
Normalized ΔF/F (sd)
2.5
0
b
Extinction without tones
only context exposure
451
'High anxiety'
'Medium anxiety'
c
Saline control injection
NaCl injection
453
No-tone extinction
Decreased act. (30%)
Same or increased act. (70%)
NaCl inj.
NaCl resp. (12%)
Unresp. (88%)
d
Tailshock
Quinine
Sucrose
All cells across all animals
630
10
e
Resp. (%) sucrose avoidance
Sucrose
Shock
Quinine
Freezing
f
5sd
30s
g
100% dF/F
10min

Supplementary Figure 10

Supporting data for 2P calcium imaging.

Related to Figure 4. **a,** Top: Visualization of the imaging periods during ExpB. The animals stayed on the setup (same context) during the entire time. Imaging periods are white boxes with duration in minutes mentioned above. Arrows indicate which fragments of the imaging sessions are shown in the four heat maps below, representing 5-minute-long neuronal fluorescence traces (rows are individual neurons) from within a single ROI in an exemplary animal undergoing ExpB. Imaging breaks during which animals stayed on the setup but the laser was shut off are marked with red two-sided arrows. Comparing the end of habituation (when anxiety is low) to the end of the fear conditioning (training, when anxiety is high) reveals a strong increase in overall activity. Note that this activity is still elevated 60 minutes later in the absence of extinction training, but rapidly decreases towards the end of extinction, when anxiety levels have decreased. All heat maps were taken from periods of the experiment when no stimuli were delivered. **b,** Top, no-tone extinction control. Animals were treated as for anxiety in Fig. 4g (see timeline above) but no tones were delivered during extinction. Heat-maps show the activity of 451 single neurons from 2 animals (rows) in the plC during 5 minutes, one hour after training (left) and 60 minutes later (right) exposed only to the context. Note, that activity upon context exposure decreases much less than when tones are used to drive extinction learning (see panel a and Fig. 4g). Bottom, while the activity of many neurons (30 %) decreased significantly, the decrease was substantially smaller than after extinction with tones, where 59 % of neurons significantly decreased their activity (tone extinction as shown in Fig. 4g). Line graph illustrates the activity decrease in the responsive population, shaded area is s.e.m. **c**, Top, saline control injection equivalent to the lithium chloride injection shown in Fig. 4g. Bottom, only 2% of neurons significantly increase their activity and to a lesser extent when saline is injected intraperitoneally. Line graph illustrates the activity decrease in the responsive population, shaded area is s.e.m. **d,** Heat maps (not binned) showing response amplitudes of all 630 single neurons recorded in all 3 animals that participated in ExpA. Neurons sorted by response strength to each of the stimuli. **e,** Proportions of responding neurons for quinine, shock and sucrose. Each dot shows the proportion of responsive cells for a single animal (n = 3 mice). **f,** Example fluorescence traces of 12 plC neurons in response to 3 tailshock presentations. Note the consistency of response for the responsive neurons. **g,** Example fluorescence traces for 4 neurons from Figure 4f containing the whole 70 min long recording without any cutouts. Note the lack of downward direction of the baseline fluorescence which would indicate bleaching/fluorescence rundown effects. Bar graphs indicate mean + s.e.m.



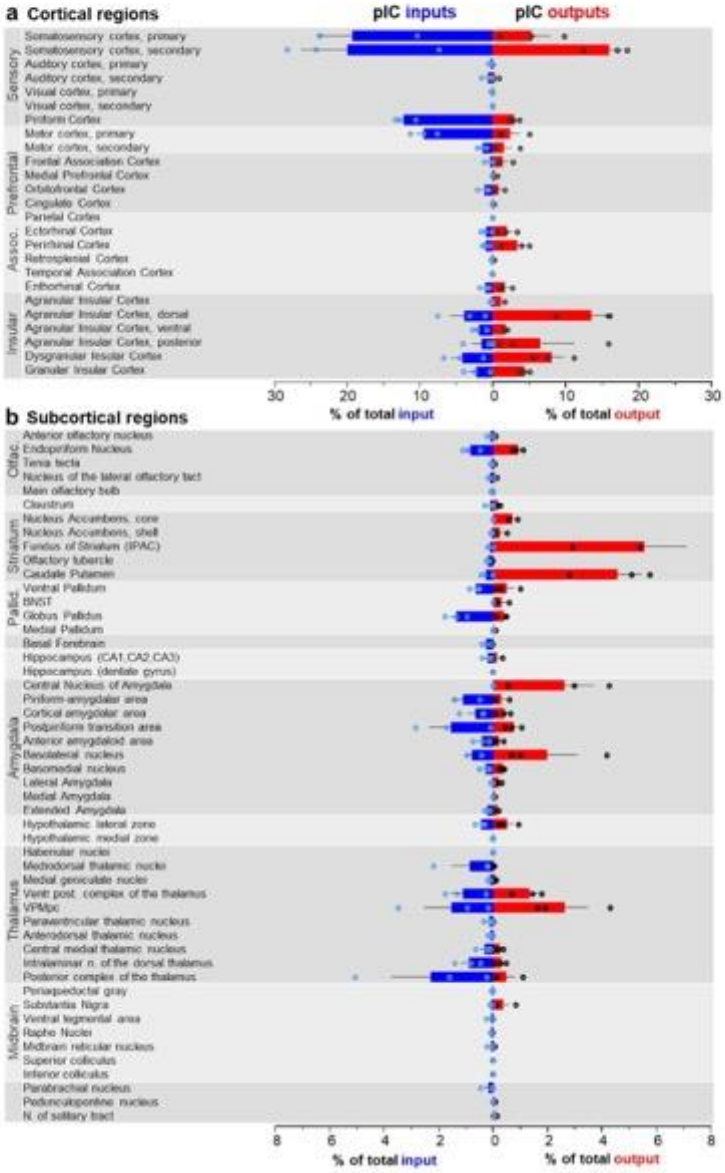
c

	# of starter cells + SD	# of input cells + SD
CamKIIa	3036.6 ± 2154.3	14400.3 ± 5504.5

Supplementary Figure 11

Starter cells of monosynaptic retrograde rabies tracings.

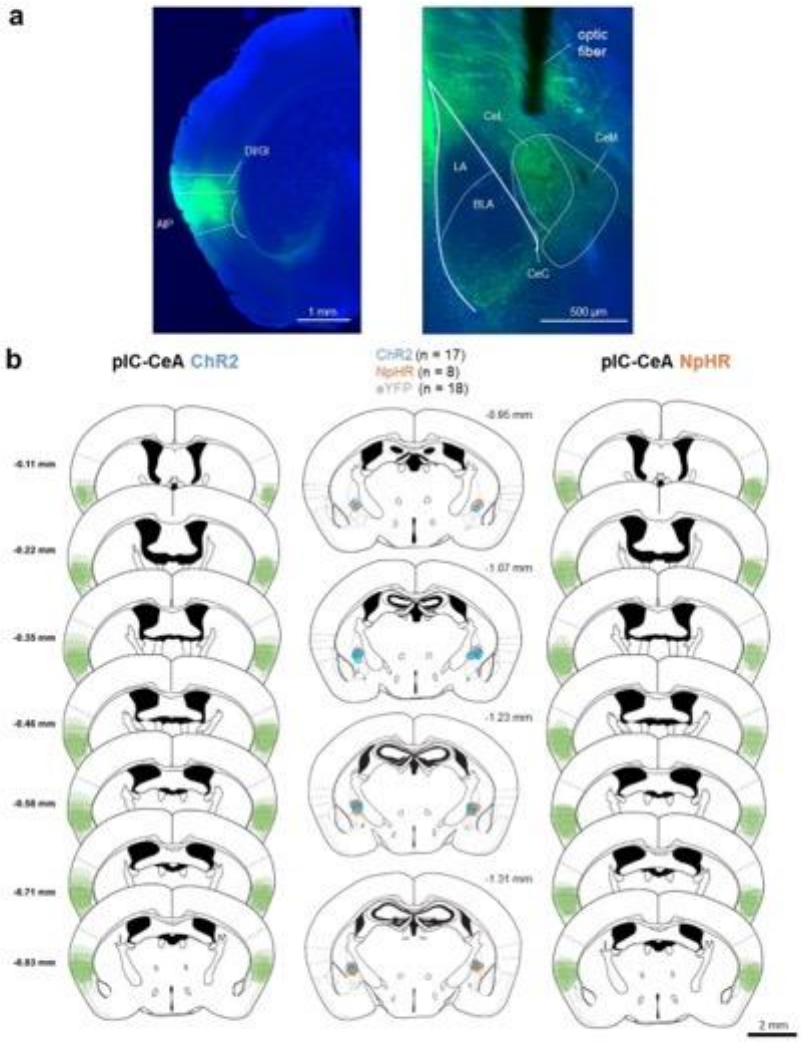
a, Confocal images of starter cells in the pIC of a CamKIIα-Cre mouse (scale bar: 100 µm). Small panels 1-3 are magnifications of starter cells highlighted with white arrowheads in the merged panel (Scale bars: 10 µm). Small panel 4 shows a TVA^+ cell that is not a starter cell, as it was not infected by rabies virus and lacks the GFP signal. Similar results were obtained in 3 independent animals. **b**, Quantification of starter cell populations. Left: Comparison of distribution of starter cells within sub-regions of the pIC. Right: Convergence index for CamKIIα-Cre rabies tracings. Data points are average cell % (left) and the convergence index (right) from n = 3 mice. **c,** Quantification of absolute starter cells and total input cells of the rabies tracings. Data are pooled from n = 3 mice. Bar graphs indicate mean + s.e.m.



Supplementary Figure 12

Whole brain input – output tracings of the pIC.

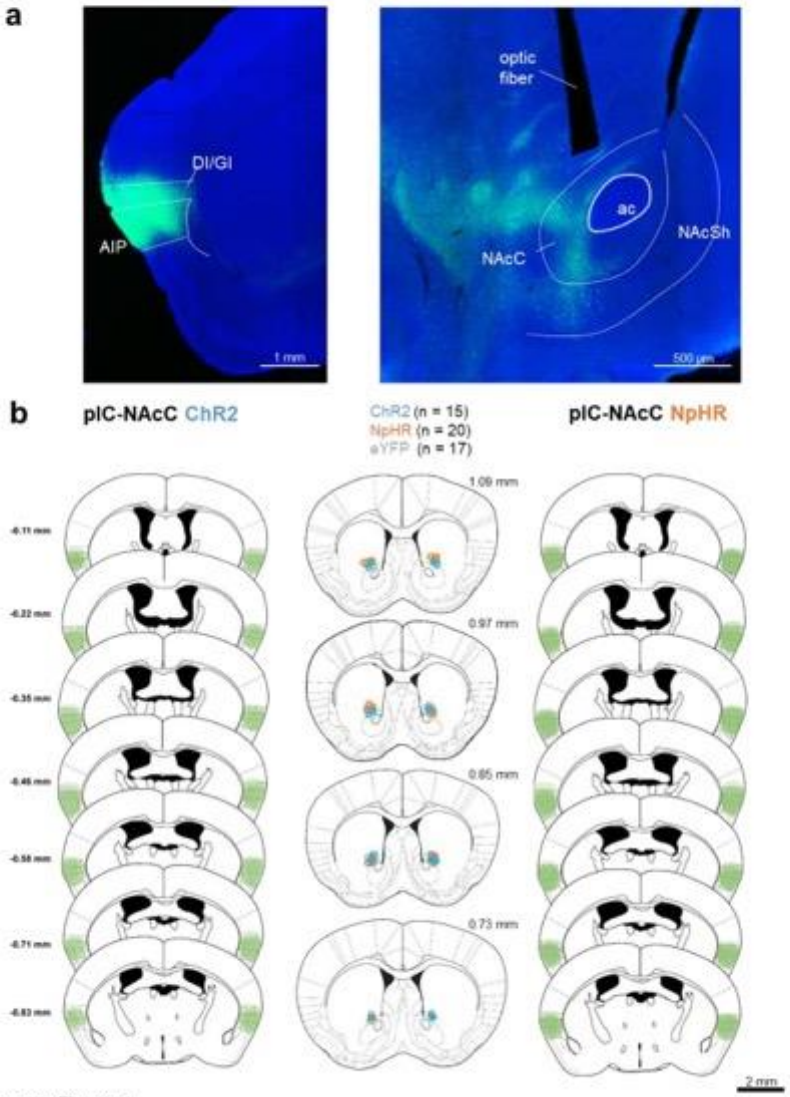
Comprehensive list of monosynaptic inputs (left) to glutamatergic (CamKIIα-Cre, in blue, $n = 3$ mice) and direct outputs from glutamatergic projection neurons (right, in red, CamKIIα-Cre $n = 3$ mice). Bar graphs indicate mean ± s.e.m.



Supplementary Figure 13

Histological verification of injection- and implantation sites of pIC-CeA-ChR2 and pIC-CeA-NpHR mice.

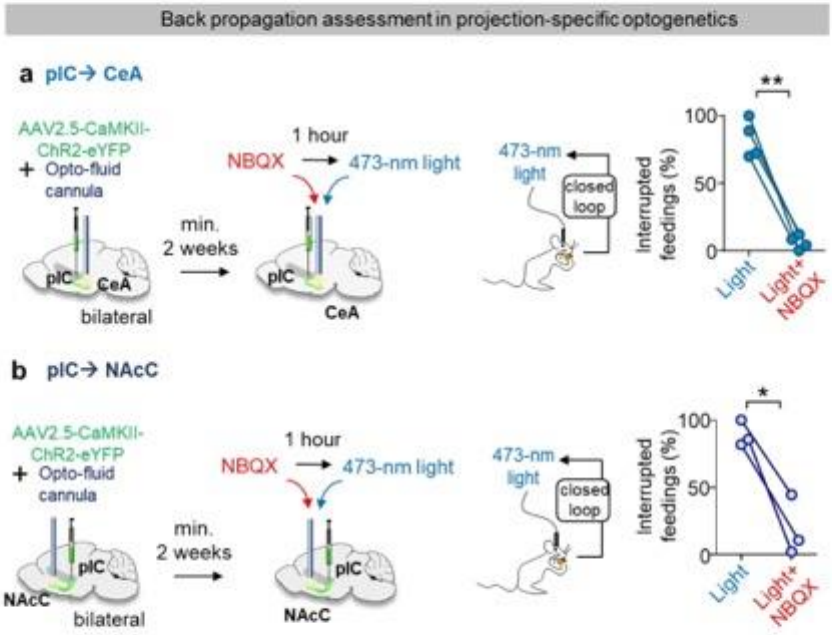
Related to Figures 7 and 8. **a**, Representative coronal sections showing the injection site (left, AP: -0.47 mm) and fiber placement (right, AP: -1.23 mm) of a pIC-CeA ChR2 animal. **b,** Left/Right: Schematic visualization of ChR2 (left) or NpHR (right) expression in the pIC. Middle: Colored dots represent the center point of the tip of the optic fibers implanted above the CeA for ChR2 (blue) or NpHR (orange) expressing animals. Implantation sites of eYFP control animals are depicted as grey circles and are partially omitted for better visibility of ChR2 and NpHR implantation sites. Pooled data from 2 independent cohorts.



Supplementary Figure 14

Histological verification of injection- and implantation sites of pIC-NAcC-ChR2 and pIC-NAcC-NpHR mice.

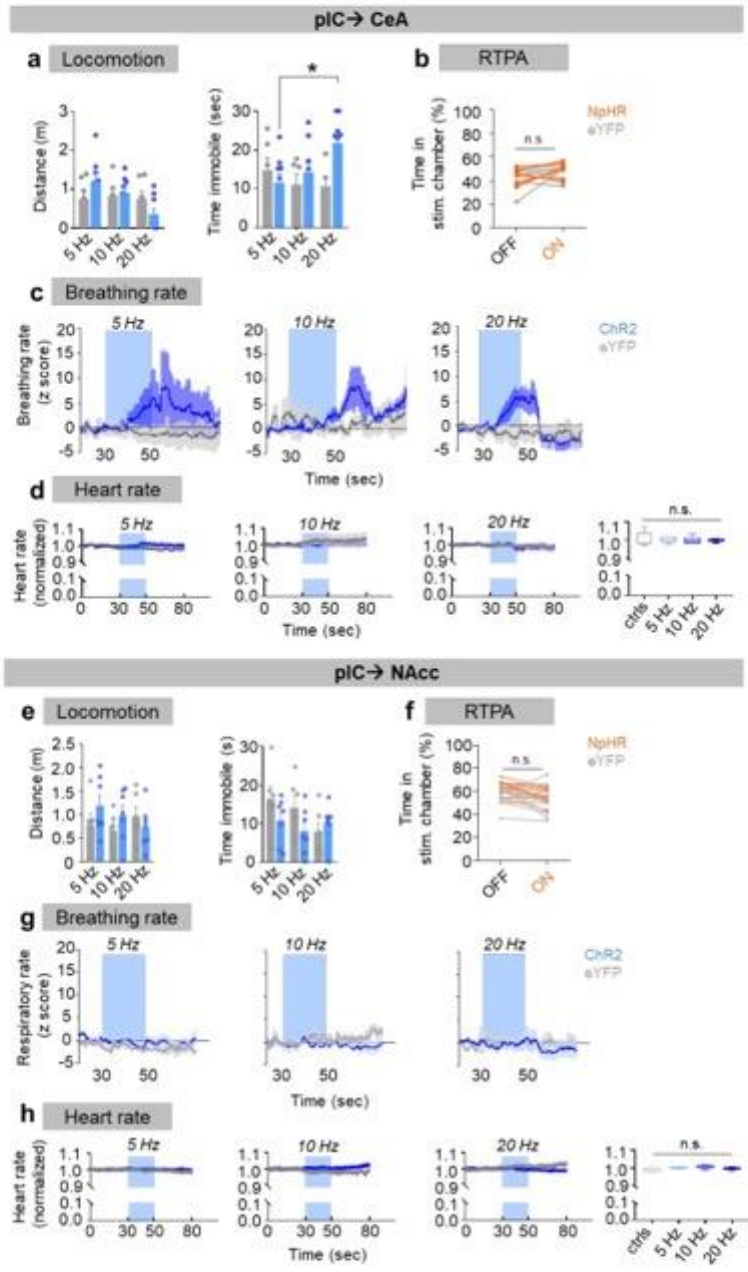
Related to Figures 7 and 8. **a**, Representative coronal sections showing the injection site (left, AP:-0.47 mm) and fiber placement (right, AP: +0.97 mm) of a pIC-NAcC ChR2 animal. **b**, Left/Right: Schematic visualization of ChR2 (left) or NpHR (right) expression in the pIC. Colored dots in the middle panel represent the center points of the optic fibers implanted above the NAcC for ChR2 (blue) or NpHR (orange) animals. Implantation sites of eYFP control animals are depicted as grey circles. Pooled data from 2 independent cohorts.



Supplementary Figure 15

Feeding interruption is not mediated via anti-dromic spikes or fibers of passage.

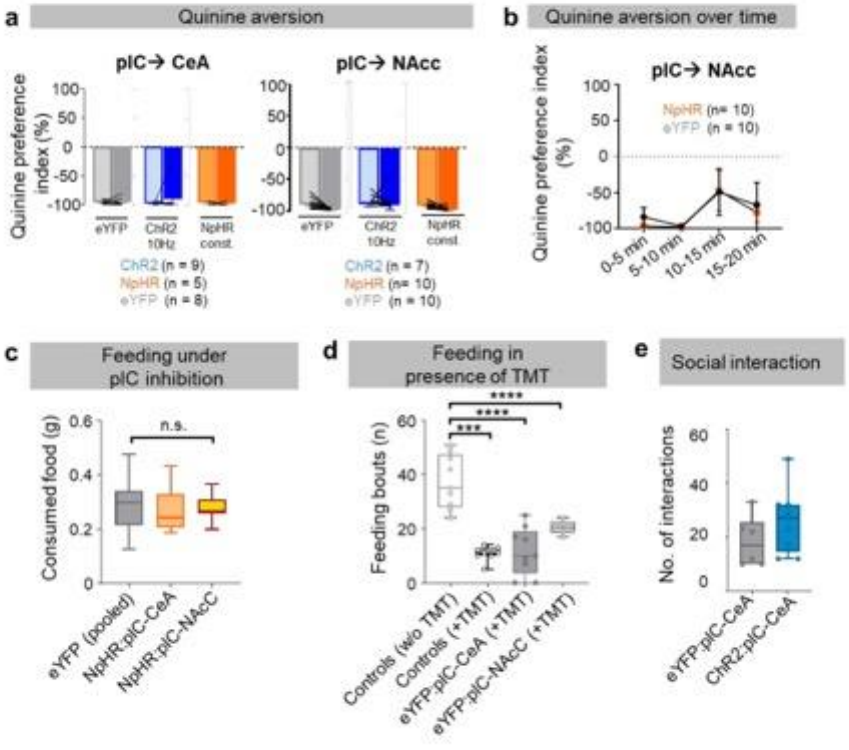
a, Two opto-fluid cannulas were bilaterally implanted over the CeA to test if feeding interruption is caused by glutamate release of insular terminals. Indeed, blocking glutamate signaling within the CeA with the AMPA-Antagonist NBQX (5 mg/ml) abolished the effect mediated by optogenetic stimulation (n = 4 mice, two-tailed paired *t* test, t = 9.162, df = 3, **p = 0.0027). **b**, We applied the same strategy for the pIC→NAcC pathway and bilaterally infused NBQX 1 h prior to optogenetic stimulation during feeding. Again, blocking AMPA-Receptors was sufficient to abolish the feeding interruption (n = 3 mice, two-tailed paired *t* test, t = 8.668, df = 2, *p = 0.0131). This suggests, that interruption of feeding is not mediated by passing fibers expressing ChR2 or by back propagation of action potentials, but by monosynaptic glutamate release from pIC terminals in the CeA and NAcC.


pIC→ CeA
a Locomotion
Distance (m)
Time immobile (sec)
5 Hz
10 Hz
20 Hz
b RTPA
Time in stim. chamber (%)
NpHR
eYFP
n.s.
OFF
ON
c Breathing rate
Breathing rate (z score)
5 Hz
10 Hz
20 Hz
ChR2
eYFP
Time (sec)
d Heart rate
Heart rate (normalized)
n.s.
ctrls
pIC→ NAcc
e Locomotion
f RTPA
g Breathing rate
Respiratory rate (z score)
h Heart rate

Supplemental Figure 16

Optogenetic manipulation of pIC→CeA, but not pIC → NAcC, influences locomotion and breathing.

Related to Figure 7. **a**, Test for frequency-dependent effects on locomotion. Activation of pIC terminals in the CeA (n = 6 eYFP mice, n = 9 ChR2 mice) did not affect the distance travelled (two-way RM ANOVA F (1, 13) = 0.07393, p = 0.7900), but activation increased the time spent immobile (two-way RM ANOVA F (1, 13) = 7.464, *p = 0.0171, Bonferroni post-hoc analysis revealed significant differences between 5 and 20 Hz in the ChR2 group *p = 0.0266). **b,** Inhibition of pIC terminals over CeA did not affect the animals' preference for the stimulated chamber (n = 7 NpHR mice, n = 6 eYFP mice, two-way RM ANOVA laser (F(1,11) = 2.959, p = 0.1134), group (F(1,11) = 0.1724, p = 0.6860).. **c,** Timeline of breathing rate responses upon 5, 10, or 20 Hz stimulations of the pIC-CeA pathway as quantified in Fig. 7i. (5 Hz: n = 4 eYFP and n = 5 ChR2 mice, 10 Hz: n = 3 eYFP and n = 3 ChR2, 20 Hz: n = 4 eYFP and n = 6 ChR2). Animals were excluded that were in a different anesthesia range. Data are z score mean ± s.e.m. **d**, Heart rate measurements upon ChR2-mediated activity manipulations of the pIC-CeA pathway. No significant effects over time or between groups were detected. Left: representative timelines from 5 Hz: n = 4 eYFP and n = 6 ChR2 mice, 10 Hz: n = 4 eYFP and n = 7 ChR2, 20 Hz: n = 4 eYFP and n = 7 ChR2 mice. Right: Averages of normalized heart-rates during the 20 s laser stimulation (n = 8 eYFP mice, n = 6 mice for 5 Hz, n = 7 mice for 10 Hz, n = 7 mice for 20 Hz; one-way ANOVA F(3,24) = 0.08556, p = 0.9673). Data is presented as min-max whisker box-plots. **e,** Activation of the pIC-NAcC pathway did not affect locomotion at any frequency (n = 6 eYFP mice, n= 7 ChR2 mice, distance: two-way RM ANOVA for frequency: F(2,48) = 2.306, p = 0.1106; and for opsin (F(3,24) = 0.4337, p = 0.7309; time immobile: two-way RM ANOVA for frequency: F (2, 22) = 1.887, p = 0.1753; and for opsin F (1, 11) = 1.592, p = 0.2331). **f,** Inhibition of pIC terminals over NAcC did not affect the animals' preference for the stimulated chamber (n = 10 mice / group, two-way RM ANOVA, opsin (F (1, 18) = 2.546, p = 0.1280). **g,** Timeline of respiratory responses upon 5, 10, or 20 Hz stimulations of the pIC-NAcC pathway in representative animals (5 Hz and 20 Hz: n = 3 eYFP and n = 6 ChR2 mice, 10 Hz: n = 3 eYFP and n = 4 ChR2). Animals were excluded that were in a different anesthesia range. **h,** Heart rate measurements upon ChR2-mediated activity manipulations of the pIC-NAcC pathway. Left: representative timelines from n = 3 eYFP and n = 6 ChR2 mice recorded in a similar anesthesia range. Right: Quantification of normalized heart-rates during the 20 s laser stimulation for 5, 10 or 20 Hz. No significant differences could be detected (n = 6 mice / group, one-way ANOVA, F(3,20) = 2.870, p = 0.0621. Bar and line graphs indicate mean + s.e.m. Box-whisker plots display median, 25th to 75th percentiles, and min to max values, as well as individual values.



Supplementary Figure 17

Supporting data for consummatory behaviors upon pIC projection-specific optogenetic manipulations.

Related to Figure 8. **a**, Optogenetic manipulation of neither the pIC-CeA (n = 9 ChR2, n = 5 NpHR, n = 8 eYFP mice) nor the pIC-NAcC (n = 7 ChR2, n = 10 NpHR, n = 10 eYFP mice) pathway impacts quinine aversion. Quinine aversion was not altered upon activation or inhibition of neither the pIC-CeA pathway (left: one-way ANOVA, $F(5, 38) = 0.4603$, $p = 0.8032$), nor the pIC-NAcC pathway (right: one-way ANOVA, $F(5, 48) = 2.117$, $p = 0.0794$). **b**, Expression of quinine aversion during the entire trial in 5 min bins. Inhibition of the pIC-NAcC pathway did not alter the course of quinine avoidance in comparison to eYFP controls (n = 10 mice / group, mixed-model ANOVA, opsin effect: $F(1, 18) = 0.1610$, $p = 0.6930$). **c,** Under naïve conditions, neither NpHR-mediated inhibition of the pIC-CeA (n=7) nor the pIC-NAcC (n=10) pathway changes feeding behavior in 24-h food deprived animals compared to controls (controls were pooled: eYFP:pIC-CeA, n=10, and eYFP:pIC-NAcC, n=10; one-way ANOVA, $F(2, 34) = 0.1531$, $p = 0.8586$). Trial duration 20 min. **d**, Feeding bouts are significantly reduced in presence of TMT. Control feeding in absence of TMT (controls w/o TMT, n = 9) was significantly more frequent than in presence of TMT (controls +TMT, n = 8), in eYFP expressing controls as shown in Fig. 8q (eYFP:pIC-CeA (+TMT), n = 9), or than in eYFP-expressing controls shown in Fig. 8r (eYFP:pIC-NAcC (+TMT), n = 6). One-way ANOVA, $F(3, 28) = 24.98$, $p < 0.0001$; Bonferroni post-hoc analyses comparing all TMT treated conditions to the no TMT control revealed significant differences (****$p < 0.0001$, ***$p = 0.0005$). **e,** Stimulation of the pIC→CeA pathway reduced the bout duration of social interactions (see Figure 8j). At the same time, the total number of interactions was slightly, but not significantly increased (n = 9 ChR2 and n = 6 eYFP mice, two-tailed unpaired *t* test, $t = 1.325$, $df = 13$, $p = 0.2081$). Bar and line graphs indicate mean + s.e.m. Box-whisker plots display median, 25th to 75th percentiles, and min to max values, as well as individual values.

5. THIRD MANUSCRIPT

From:

Dolensek, N., & Gogolla, N. (2020). Machine-learning approaches to classify and understand emotion states in mice. *Neuropsychopharmacology 2020 46:1*. https://www.nature.com/articles/s41386-020-00857-8

www.nature.com/npp

Neuropsychopharmacology

HOT TOPICS

Machine-learning approaches to classify and understand emotion states in mice

Nejc Dolensek[1,2] and Nadine Gogolla[1]

Neuropsychopharmacology (2021) 46:250–251; https://doi.org/10.1038/s41386-020-00857-8

Emotions are one of the most intriguing products of brain function, yet, we still lack a mechanistic understanding of how emotions arise in the brain [1]. One of the major hindrances to studying the neuronal mechanisms of emotion lies in the difficulty to link observable behavior with internal states and ongoing neural activity [1].

In a recent study, we employed machine-vision and -learning approaches to investigate facial expressions of emotions and their neuronal correlates in mice [2]. Our work built on prior evidence that mice use their orofacial musculature to react to sensory stimuli such as tastants [3] or pain [4]. To assess facial expressions objectively and to detect previously uncharacterized changes, we recorded close-up videos of head-fixed mice and extracted features from video frames using "histograms of oriented gradients" (HOG), a machine-vision technique that represents the statistics of local image features. This allowed us to compare facial expressions of mice reacting to emotion events quantitatively in an unsupervised manner using hierarchical clustering. We next asked whether we could separate facial expressions exhibited upon diverse emotionally salient events into distinct categories by reducing the dimensionality of our data via principal component analysis (PCA) followed by t-distributed stochastic neighbor embedding (t-SNE). This approach separated facial expressions into discrete emotions.

Further, we were able to train a random forest classifier, a supervised machine-learning algorithm, to distinguish different facial expressions across mice with >90% accuracy. We next created "prototypical faces" which enabled us to resolve the intensity of each emotion at millisecond timescales and confirmed that facial expressions corresponded to emotional states, rather than reflex-like reactions. Indeed, exploiting the quantitative nature of our machine-vision approach allowed us to demonstrate that facial expressions revealed core properties of emotion such as intensity, persistence, flexibility, and valence [1]. Furthermore, the same facial expressions that resulted from external sensory triggers were also evoked by optogenetic manipulations in emotion-relevant brain circuits. Finally, we aligned facial tracking with neural activity recordings via two-photon calcium imaging and identified single neurons in the insular cortex whose activity closely correlated with specific emotional facial expressions.

Our results establish a framework for quantitative and objective assessments of distinct emotion features and categories in mice, one of the most prevalent model organisms in neuroscience. Machine-learning approaches, as utilized in our study, hold promise to unravel the neuronal underpinnings of emotion processes and gain better definition of emotion states since they can identify and describe previously uncharacterized emotion-related changes at high temporal resolution. The use of such quantitative assessments of affective states further opens new doors to psychopharmacological research investigating how substances affect emotional states. Insights into the nature and mechanisms of emotion processing are of uttermost importance for the clinic, since many psychiatric disorders involve emotional dysfunction.

While facial expressions constitute only one aspect of affective states, the current surge of supervised and unsupervised machine-learning approaches [5, 6] presents an unprecedented opportunity to create multi-dimensional models of affect through a combination of accurately tracked changes in animal posture, behavioral patterns and physiological changes along with large-scale neural recordings. While the relationships between these distinct parameters are complex, artificial intelligence approaches hold promise for reducing this high dimensionality and thus overcome current limitations in creating mechanistic hypotheses of how the brain generates emotion.

FUNDING AND DISCLOSURE

This work was supported by the Max Planck Society and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC2017-STG, grant agreement no. 758448 to N.G.). Unfortunately, we are unable to comprehensively cite all relevant literature due to space limitations and would thus like to thank colleagues across the fields of ethology, computational biology, artificial intelligence and affective neuroscience for inspiration. The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

ND and NG wrote this manuscript.

ADDITIONAL INFORMATION

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Anderson DJ, Adolphs R. A framework for studying emotions across species. Cell 2014;157:187–200.

[1]Circuits for Emotion Research Group, Max Planck Institute of Neurobiology, Am Klopferspitz 18, 82152 Martinsried, Germany and [2]Graduate School of Systemic Neurosciences, Ludwig-Maximilians University Munich, Munich, Germany
Correspondence: Nadine Gogolla (ngogolla@neuro.mpg.de)

Published online: 8 September 2020

2. Dolensek N, Gehrlach DA, Klein AS, Gogolla N. Facial expressions of emotion states and their neuronal correlates in mice. Science. 2020;368:89–94.
3. Grill HJ, Norgren R. The taste reactivity test. I. Mimetic responses to gustatory stimuli in neurologically normal rats. Brain Res. 1978;143:263–79.
4. Langford DJ, et al. Coding of facial expressions of pain in the laboratory mouse. Nat Methods 2010;7:447–9.
5. Goodwin NL, Nilsson SRO, Golden SA. Rage against the machine: advancing the study of aggression ethology via machine learning. Psychopharmacology. 2020. https://doi.org/10.1007/s00213-020-05577-x.
6. Datta SR, Anderson DJ, Branson K, Perona P, Leifer A. Computational neuroethology: a call to action. Neuron 2019;104:11–24.

6. DISCUSSION

Each of the three respective manuscripts, which form the heart of this dissertation, already contains a detailed discussion of its findings standing on its own. Therefore, I am going to use this section to focus on the findings, to which I solely significantly contributed, attempt to synthesize them, and discuss how they fit in the wider field of emotion research.

The main findings of my work presented in this thesis are as follows: first, mice exhibit stereotyped facial expressions of emotion, which do not depend on, or simply reflect, responses to physical stimuli, but rather echo internal emotion states. Second, these facial expressions of emotion can be automatically and objectively detected and quantitatively measured using machine vision approaches I developed, resulting in a moment-to-moment description. Third, insular cortex, an understudied brain region known to be strongly involved in emotion processing, encodes emotionally salient stimuli across modalities. Fourth, while some single neurons in insular cortex displayed tightly stimulus-locked responses, much of the neural activity in this region could not easily be explained by sensory stimulus presentation, because of its duration, timing, and magnitude, indicating encoding of some other process. Fifth and last, emotional state readouts as derived from facial expressions strongly correlate with activity of single neurons in insular cortex, a brain region known to be strongly involved in emotion, and uncovered distinct neural subpopulations associated with different emotions.

My work on facial expressions of emotion and their neural code introduces tools and findings that make emotions tangible in mice and enables deep dissection of neural circuits involved in emotion using advanced modern neuroscientific tools available to mouse neuroscientists. In addition to that it further supports mice as a valid model organism for studying emotion, with their emotional experience seemingly not unlike that of other mammals or even humans, which has long been a point of heated contention. However, a traditional view on facial expressions sees them as mainly having a social role. While recent work has shown that rodents do experience social transfer of emotion to some degree (Ferretti & Papaleo, 2018), it has long been known that mice rely on a plethora of olfactory and tactile cues to communicate,

with facial expressions serving a yet unknown role. Nonetheless, many mammals, including rodents, do interact using their faces (Brecht & Freiwald, 2012). Leaning on existing literature and my findings, I propose a potential triple function of facial expressions of emotion, not only in mice but also in other organisms known to exhibit them, like primates. The three functions I hypothesize are the defensive, social and feedback roles.

Defensive role has already been proposed by Darwin (Darwin & Darwin, 2009) and to some degree by Grill, Norgren and Berridge (Berridge & Kringelbach, 2013; Grill & Norgren, 1978a). In this role, facial musculature either exposes, to enhance sensory acquisition, or protects sensory and vital organs when appetitive or aversive stimuli are presented. For example, a mouse, but also a monkey, or a baby will retract its tongue and close its mouth immediately after tasting a bitter substance, preventing further stimulation of taste buds, and blocking the entry of the potentially dangerous substance. On another hand, tasting a sweet tastant will make an animal extend its tongue and open its mouth, exposing itself to more of the substance. This might indicate that facial expressions are rather specific to the modality of the stimulus and selectively block or expose the sensory organs directly exposed to a specific stimulus. Crucially, I observed that not to be the case. A painful electric shock delivered to a mouse's tail, as far away from the face as possible, caused a profound change in facial expression, with eyes closing and ears being pressed to the body even though they received no sensory stimulation of any kind. I observed that all aversive stimuli resulted in facial expressions protecting most sensory organs of the face, regardless of the modality or location of the stimulation. On another hand appetitive stimuli caused facial expressions exposing these same sensory organs across the board. A study (Susskind et al., 2008) performed on humans reported a related finding, where exhibiting facial expressions of fear or disgust had an effect on the nasal volume and the amount of air inhaled, with both increasing with a fearful expression (leading to enhanced sensory acquisition in a potentially dangerous situation) and decreasing with a disgusted expression (leading to reduced sensory acquisition when in contact with a potentially noxious substance). This might indicate that facial expressions serve a sort of a heuristic adaptive and protective role, where a stimulus is almost immediately assigned to be associated with one of a handful of "broader" basic emotions, triggering a set of less-than-specific orofacial musculature contractions.

The second role of facial expressions might be the **social role**. Facial expressions in humans are known to serve an important role in social interaction, by providing a readout of one's emotional state to people nearby. Ability to perceive otherwise hidden internal state of actors in your surroundings is certainly an evolutionary beneficial trait and has been broadly observed. Animals, including rodents, were shown to be able to use visual information to detect emotion in conspecifics and even other species (Albuquerque et al., 2016; Bartal et al., 2011; Ferretti & Papaleo, 2018; Mason, 2021). However, it has yet to be shown whether mice can perceive emotional expressions of one another. An interesting initial approach could be to measure whether mice observe facial features of other mice using eye tracking.

The third and final role of facial expressions of emotion might be the **feedback role**. Bodily expressions of emotion have long been proposed to be significant components of emotion, with James-Lange theory of emotion and Facial Feedback Hypothesis even postulating a central role (Zych & Gogolla, 2021). Many studies support this position at least to some degree, with a series of studies showing that facial Botox injections (Vianna et al., 2006), which impair the movement of facial muscles, result in a diminished emotional experience and a reduction in activation of brain regions like amygdala (Kim et al., 2014) after emotional stimulus presentation. A recent literature metanalysis (Coles et al., 2019) reported that effects of facial feedback in humans are significant, but small. Studies in rodents have shown that especially breathing and freezing have a strong effect on emotion (Bagur et al., 2018). Curiously, orofacial movements in resting mice were found to have a significant influence on neural activity in visual cortex (Stringer et al., 2019), indicating a perhaps global feedback effect.

Importantly, it is possible and probable that facial expressions don't serve a singular role but multiple simultaneously. More research work, systematically testing each of the roles described above is necessary to gain a more complete picture. An especially interesting approach to answering these questions would be to compare neural mechanisms for facial expressions to those known to drive other better studied emotional behaviors, for example freezing, escapes, approaches, and avoidance. Facial expressions might be revealing otherwise hidden (as in not considered to be directly resulting in outwardly observable behavior) cognitive neural processes, which could be separate in function and neural implementation from the neural circuitry driving

classical emotional behaviors. Emotions could be a form of intermediate neural representations, created in the brain on the path from sensory perception to action, serving mainly to rapidly modulate both the upstream and the downstream processing, both via direct neural connections, but also indirectly, by for example exerting influence on the orofacial musculature, where many sensory organs are located. Insular cortex is an interesting candidate for a such function, considering its strong bidirectional connectivity with sensory input regions like the thalamus and brain regions known to drive emotional behaviors like the amygdalar and brain-stem nuclei. Furthermore, studies (Gehrlach et al., 2019; Livneh et al., 2017) have shown that insula, while having a profound effect on a wide range of emotional behaviors, likely mainly serves a gating or biasing role for its downstream regions, which are actually necessary and sufficient to produce behavior. On another hand, insula is possibly necessary for production of facial expressions of emotion (Langford et al., 2010; Shepherd & Freiwald, 2018) and their perception (Lin et al., 2016; Phillips et al., 1997; Terasawa et al., 2015), likely via a form of mirror neurons. Same circuitry implementing emotional processes themselves, their behavioral indicators in the form of facial expressions and also their perception fits models long discussed by empathy researchers (Iacoboni & Lenzi, 2002) and elegantly fits existing data, including my findings.

My analysis of facial expressions suggested they are discrete, and only continuous intensity-wise for each single emotion. This was further supported by the finding of non-overlapping "face" neurons in insular cortex, activity of which correlated with only a single facial expression. This is an especially interesting finding, since it relates to one of the oldest questions in the science of emotion – do emotions exist on a spectrum or are they discrete categories? However, neural activity of only two emotions was tested in the first manuscript. A future study would require the use of a battery of stimuli used in the calcium imaging experiments in the second manuscript, which resulted in strong insular cortex responses, find "face" neurons for each emotion, and confirm that the "face" neurons across all emotions are still non-overlapping. Likewise, the discreteness of facial expressions might be related to the analysis approach used. Principal component embedding indicated that all facial expressions can be represented on two to three common axes with only some loss of information and thus more unsupervised analysis approaches might be needed to answer this question. Therefore, evidence is still inconclusive whether emotions are discrete or continuous.

Finally, I would like to discuss how my work fits into the broader field of emotion research and cover some potential pitfalls. Facial expressions are interesting in their own right but it is important to note that it would be a mistake to deduce that emotional experience can be completely reduced to expression. While expression might reflect and even significantly affect neural states related to emotion it likely only reveals a platonic shadow of complex changes in activity distributed across large parts of the nervous system. Even behavior-wise it is not likely that facial expressions stand alone, even though they do offer some unique advantages over the more established behavioral readouts of emotion states. For example, optogenetic perturbations that elicited facial expressions in manuscript 1 also elicited many other more commonly studied behaviors, for example avoidance and approach behaviors and changes in pupil size, as covered in manuscript 2. Furthermore, in a classic study, Grill and Norgren (1978) demonstrated that even chronically decerebrate rats exhibit some orofacial movements and motor behaviors like rearing in response delivery of different tastants, indicating that while emotional behaviors can be driven by activity in high-level brain regions associated with subjective emotional experiences like insula, these behaviors are not necessarily specific to activity in those regions. In addition to that, any behavioral measurement is unavoidably at least a step removed from any neural activity driving it as it relies on observing results of muscular contractions which do not necessarily have the same dynamics as the neural systems driving them. Instead of suggesting that emotion can be completely reduced to its behavioral correlates, I believe that these instead offer excellent entry points to a rigorous study of the neural circuitry in which emotion is actually manifested. The fact that both the expressions themselves and the related emotional activity in insular cortex recorded reflected many proposed qualities of emotion (Anderson & Adolphs, 2014) is a confirmation of this belief.

Taken together, research presented in this thesis presents a significant step forward in the study of emotions by enabling the use of a rodent model, with all accompanying tools, to research the neuronal basis of emotion, function of insular cortex and the role of facial expressions.

7. REFERENCES

Albuquerque, N., Guo, K., Wilkinson, A., Savalli, C., Otta, E., & Mills, D. (2016). Dogs recognize dog and human emotions. *Biology Letters*, *12*(1). https://doi.org/10.1098/rsbl.2015.0883

Anderson, D. J., & Adolphs, R. (2014). A framework for studying emotions across species. *Cell*, *157*(1), 187–200. https://doi.org/10.1016/j.cell.2014.03.003

Bagur, S., Lefort, J. M., Lacroix, M. M., de Lavilléon, G., Herry, C., Billand, C., Geoffroy, H., & Benchenane, K. (2018). Dissociation of fear initiation and maintenance by breathing-driven prefrontal oscillations. In *bioRxiv*. bioRxiv. https://doi.org/10.1101/468264

Bartal, I. B. A., Decety, J., & Mason, P. (2011). Empathy and pro-social behavior in rats. *Science*, *334*(6061), 1427–1430. https://doi.org/10.1126/science.1210789

Berridge, K. C. (2018). *Evolving Concepts of Emotion and Motivation*. https://doi.org/10.3389/fpsyg.2018.01647

Berridge, K. C., & Kringelbach, M. L. (2013). Neuroscience of affect: Brain mechanisms of pleasure and displeasure. In *Current Opinion in Neurobiology* (Vol. 23, Issue 3, pp. 294–303). https://doi.org/10.1016/j.conb.2013.01.017

Brecht, M., & Freiwald, W. A. (2012). The many facets of facial interactions in mammals. In *Current Opinion in Neurobiology* (Vol. 22, Issue 2, pp. 259–266). Elsevier Current Trends. https://doi.org/10.1016/j.conb.2011.12.003

Caeiro, C. C., Burrows, A. M., & Waller, B. M. (2017). Development and application of CatFACS: Are human cat adopters influenced by cat facial expressions? *Applied Animal Behaviour Science*, *189*, 66–78. https://doi.org/10.1016/j.applanim.2017.01.005

Coles, N. A., Larsen, J. T., & Lench, H. C. (2019). *A Meta-Analysis of the Facial Feedback Literature: Effects of Facial Feedback on Emotional Experience Are Small and Variable*. https://doi.org/10.1037/bul0000194

Darwin, C., & Darwin, F. (2009). The expression of the emotions in man and animals. In *The Expression of the Emotions in Man and Animals*. https://doi.org/10.1017/CBO9780511694110

De La Torre, F., Simon, T., Ambadar, Z., & Cohn, J. F. (2011). Fast-FACS: A computer-assisted system to increase speed and reliability of manual FACS coding. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6974 LNCS*(PART 1), 57–66. https://doi.org/10.1007/978-3-642-24600-5_9

Defensor, E. B., Corley, M. J., Blanchard, R. J., & Blanchard, D. C. (2012). Facial expressions of mice in aggressive and fearful contexts. *Physiology and Behavior*, *107*(5), 680–685. https://doi.org/10.1016/j.physbeh.2012.03.024

Ekman, P., & Friesen, W. (1978). Facial action coding system: a technique for the measurement of facial movement. *Palo Alto: Consulting Psychologists Press.*

Ekman, Paul, Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86–88. https://doi.org/10.1126/science.164.3875.86

Ferretti, V., & Papaleo, F. (2018). Understanding others: emotion recognition abilities in humans and other animals. *Genes, Brain and Behavior*, *18*(1), e12544. https://doi.org/10.1111/gbb.12544

Charles le Brun, The Expressions.jpg - Wikimedia Commons. (2011, June 8). https://commons.wikimedia.org/wiki/File:Charles_le_Brun,_The_Expressions.jpg

Finlayson, K., Lampe, J. F., Hintze, S., Würbel, H., & Melotti, L. (2016). Facial indicators of positive emotions in rats. *PLoS ONE*, *11*(11), e0166446. https://doi.org/10.1371/journal.pone.0166446

Gehrlach, D. A., Dolensek, N., Klein, A. S., Roy Chowdhury, R., Matthys, A., Junghänel, M., Gaitanos, T. N., Podgornik, A., Black, T. D., Reddy Vaka, N., Conzelmann, K. K., & Gogolla, N. (2019). Aversive state processing in the posterior insular cortex. *Nature Neuroscience*, *22*(9), 1424–1437. https://doi.org/10.1038/s41593-019-0469-1

Gehrlach, D. A., Weiand, C., Gaitanos, T. N., Cho, E., Klein, A. S., Hennrich, A. A., Conzelmann, K. K., & Gogolla, N. (2020). A whole-brain connectivity map of mouse insular cortex. *ELife*, *9*, 1–78. https://doi.org/10.7554/ELIFE.55585

Gogolla, N. (2017). The insular cortex. In *Current Biology* (Vol. 27, Issue 12, pp. R580–R586). Cell Press. https://doi.org/10.1016/j.cub.2017.05.010

Grill, H. J., & Norgren, R. (1978a). The taste reactivity test. I. Mimetic responses to gustatory stimuli in neurologically normal rats. *Brain Research*, *143*(2), 263–279. https://doi.org/10.1016/0006-8993(78)90568-1

Grill, H. J., & Norgren, R. (1978b). Chronically decerebrate rats demonstrate satiation but not bait shyness. *Science*, *201*(4352), 267–269. https://doi.org/10.1126/science.663655

Guillaume Duchenne de Boulogne performing facial electrostimulus experiments.jpg - Wikimedia Commons. (2005, September 1). https://commons.wikimedia.org/wiki/File:Guillaume_Duchenne_de_Boulogne_performing_facial_electrostimulus_experiments.jpg

Hartley, L. (2001). *Physiognomy and the Meaning of Expression in Nineteenth-Century Culture*. Cambridge University Press.

Hueston, J. T., & Cuthbertson, R. A. (1978). Duchenne de Boulogne and facial expression. *Annals of Plastic Surgery*, *1*(4), 411–420. https://doi.org/10.1097/00000637-197807000-00009

Iacoboni, M., & Lenzi, G. L. (2002). Mirror neurons, the insula, and empathy. In *Behavioral and Brain Sciences* (Vol. 25, Issue 1, pp. 39–40). Cambridge University Press. https://doi.org/10.1017/S0140525X02420018

Ikenoue, E., Akhter, F., Tsutsumi, Y., Sato, F., Ohara, H., Uchino, K., Furuta, T., Tachibana, Y., & Yoshida, A. (2018). Transcortical descending pathways through granular insular cortex conveying orofacial proprioception. *Brain Research*, *1687*, 11–19. https://doi.org/10.1016/j.brainres.2018.02.033

Kim, M. J., Neta, M., Davis, F. C., Ruberry, E. J., Dinescu, D., Heatherton, T. F., Stotland, M. A., & Whalen, P. J. (2014). Botulinum toxin-induced facial muscle paralysis affects amygdala responses to the perception of emotional expressions: Preliminary findings from an A-B-A design. *Biology of Mood and Anxiety Disorders*, *4*(1), 11. https://doi.org/10.1186/2045-5380-4-11

Kupfermann, I., & Weiss, K. (1978). The command neuron concept. Behavioral and Brain Sciences, 1(1), 3-10. doi:10.1017/S0140525X00059057

Langford, D. J., Bailey, A. L., Chanda, M. L., Clarke, S. E., Drummond, T. E., Echols, S., Glick, S., Ingrao, J., Klassen-Ross, T., Lacroix-Fralish, M. L., Matsumiya, L., Sorge, R. E., Sotocinal, S. G., Tabaka, J. M., Wong, D., Van Den Maagdenberg, A. M. J. M., Ferrari, M. D., Craig, K. D., & Mogil, J. S. (2010). Coding of facial expressions of pain in the

laboratory mouse. *Nature Methods*, *7*(6), 447–449. https://doi.org/10.1038/nmeth.1455
Lin, H., Mueller-Bardorff, M., Mothes-Lasch, M., Buff, C., Brinkmann, L., Miltner, W. H. R., & Straube, T. (2016). Effects of Intensity of Facial Expressions on Amygdalar Activation Independently of Valence. *Frontiers in Human Neuroscience*, *10*(DEC2016), 646. https://doi.org/10.3389/fnhum.2016.00646
Livneh, Y., Ramesh, R. N., Burgess, C. R., Levandowski, K. M., Madara, J. C., Fenselau, H., Goldey, G. J., Diaz, V. E., Jikomes, N., Resch, J. M., Lowell, B. B., & Andermann, M. L. (2017). Homeostatic circuits selectively gate food cue responses in insular cortex. *Nature*, *546*(7660), 611–616. https://doi.org/10.1038/nature22375
Lyons, J. D. (2019). The Oxford Handbook of the Baroque. In *The Oxford Handbook of the Baroque*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190678449.001.0001
Mason, P. (2021). Lessons from helping behavior in rats. In *Current Opinion in Neurobiology* (Vol. 68, pp. 52–56). Elsevier Ltd. https://doi.org/10.1016/j.conb.2021.01.001
Meineck, P. (2011). The neuroscience of the tragic mask. In *Arion - Journal of Humanities and the Classics* (Vol. 19, Issue 1, pp. 113–158). https://doi.org/10.2307/41308596
Méndez-Ruette, M., Linsambarth, S., Moraga-Amaro, R., Quintana-Donoso, D., Méndez, L., Tamburini, G., Cornejo, F., Torres, R. F., & Stehberg, J. (2019). The Role of the Rodent Insula in Anxiety. *Frontiers in Physiology*, *10*(MAR), 330. https://doi.org/10.3389/fphys.2019.00330
Micheletta, J., Whitehouse, J., Parr, L. A., & Waller, B. M. (2015). Facial expression recognition in crested macaques (Macaca nigra). *Animal Cognition*, *18*(4), 985–990. https://doi.org/10.1007/s10071-015-0867-z
Panksepp, J. (2011). The basic emotional circuits of mammalian brains: Do animals have affective lives? In *Neuroscience and Biobehavioral Reviews* (Vol. 35, Issue 9, pp. 1791–1804). https://doi.org/10.1016/j.neubiorev.2011.08.003
Parent, A. (2005). Duchenne De Boulogne: a pioneer in neurology and medical photography. *The Canadian Journal of Neurological Sciences. Le Journal Canadien Des Sciences Neurologiques*, *32*(3), 369–377. https://doi.org/10.1017/S0317167100004315
Parr, L. A., Waller, B. M., Burrows, A. M., Gothard, K. M., & Vick, S. J. (2010). Brief communication: MaqFACS: A muscle-based facial movement coding system for the rhesus macaque. *American Journal of Physical Anthropology*, *143*(4), 625–630. https://doi.org/10.1002/ajpa.21401
Parr, Lisa A., Waller, B. M., & Fugate, J. (2005). Emotional communication in primates: Implications for neurobiology. In *Current Opinion in Neurobiology* (Vol. 15, Issue 6, pp. 716–720). https://doi.org/10.1016/j.conb.2005.10.017
Parr, Lisa A., Waller, B. M., Vick, S. J., & Bard, K. A. (2007). Classifying chimpanzee facial expressions using muscle action. *Emotion*, *7*(1), 172–181. https://doi.org/10.1037/1528-3542.7.1.172
Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. In *Brain structure & function* (Vol. 214, Issues 5–6, pp. 451–463). https://doi.org/10.1007/s00429-010-0258-9
Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., Bullmore, E. T., Perrett, D. I., Rowland, D., Williams, S. C. R., Gray, J. A., & David, A. S. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*(6650), 495–498. https://doi.org/10.1038/39051
Samadiani, N., Huang, G., Cai, B., Luo, W., Chi, C. H., Xiang, Y., & He, J. (2019). A review on

automatic facial expression recognition systems assisted by multimodal sensor data. In *Sensors (Switzerland)* (Vol. 19, Issue 8). MDPI AG. https://doi.org/10.3390/s19081863

Sato, F., Uemura, Y., Kanno, C., Tsutsumi, Y., Tomita, A., Oka, A., Kato, T., Uchino, K., Murakami, J., Haque, T., Tachibana, Y., & Yoshida, A. (2017). Thalamo-insular pathway conveying orofacial muscle proprioception in the rat. *Neuroscience*, *365*, 158–178. https://doi.org/10.1016/j.neuroscience.2017.09.050

Shepherd, S. V., & Freiwald, W. A. (2018). Functional Networks for Social Communication in the Macaque Monkey. *Neuron*, *99*(2), 413-420.e3. https://doi.org/10.1016/j.neuron.2018.06.027

Sliz, D., & Hayley, S. (2012). Major depressive disorder and alterations in insular cortical activity: A review of current functional magnetic imaging (fMRI) research. In *Frontiers in Human Neuroscience* (Vol. 6, Issue NOVEMBER 2012). Frontiers Media S. A. https://doi.org/10.3389/fnhum.2012.00323

Sotocinal, S. G., Sorge, R. E., Zaloum, A., Tuttle, A. H., Martin, L. J., Wieskopf, J. S., Mapplebeck, J. C. S., Wei, P., Zhan, S., Zhang, S., McDougall, J. J., King, O. D., & Mogil, J. S. (2011). The Rat Grimace Scale: A partially automated method for quantifying pain in the laboratory rat via facial expressions. *Molecular Pain*, *7*, 1744-8069-7–55. https://doi.org/10.1186/1744-8069-7-55

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., & Harris, K. D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, *364*(6437). https://doi.org/10.1126/science.aav7893

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience*, *11*(7), 843–850. https://doi.org/10.1038/nn.2138

Terasawa, Y., Kurosaki, Y., Ibata, Y., Moriguchi, Y., & Umeda, S. (2015). Attenuated sensitivity to the emotions of others by insular lesion. *Frontiers in Psychology*, *6*, 1314. https://doi.org/10.3389/fpsyg.2015.01314

Tsutsumi, Y., Tachibana, Y., Sato, F., Furuta, T., Ohara, H., Tomita, A., Fujita, M., Moritani, M., & Yoshida, A. (2018). Cortical and Subcortical Projections from Granular Insular Cortex Receiving Orofacial Proprioception. *Neuroscience*, *388*, 317–329. https://doi.org/10.1016/j.neuroscience.2018.07.047

Vianna, E. P. M., Weinstock, J., Elliott, D., Summers, R., & Tranel, D. (2006). Increased feelings with increased body signals. *Social Cognitive and Affective Neuroscience*, *1*(1), 37–48. https://doi.org/10.1093/scan/nsl005

Waller, B. M., Peirce, K., Caeiro, C. C., Scheider, L., Burrows, A. M., McCune, S., & Kaminski, J. (2013). Paedomorphic facial expressions give dogs a selective advantage. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0082686

Wathan, J., Burrows, A. M., Waller, B. M., & McComb, K. (2015). EquiFACS: The equine facial action coding system. *PLoS ONE*, *10*(8). https://doi.org/10.1371/journal.pone.0131738

Winkielman, P., Berridge, K. C., & Wilbarger, J. L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin*, *31*(1), 121–135. https://doi.org/10.1177/0146167204271309

Zych, A. D., & Gogolla, N. (2021). Expressions of emotions across species. *Current Opinion in Neurobiology*, *68*, 57–66. https://doi.org/10.1016/j.conb.2021.01.003

8. MISCELLANEOUS

8.1 ACKNOWLEDGEMENTS

I have always wanted to become a scientist and you could say that the last 5 years were truly a part of a dream come true and helped make me as excited about science as ever. These years were just as challenging, stimulating, exhilarating and fun as I imagined, and they would not have been the same without the people mentioned below. I will forever be thankful for that.

Firstly, I want to thank Nadine Gogolla, to whom I am beyond grateful for giving me the opportunity to work in her lab at its very start. I can confidently say that this experience has shaped me as a scientist and the many lessons learned will stay with me forever. It wasn't always easy, but I would choose the same path again, every single time.

Next, I'd like to thank my thesis advisory committee members Mark Hübener and Anton Sirota, but also Ruben Portugues for their invaluable input on my scientific work, but also personal and career advice.

Following that, I'd like to acknowledge my lab-mates for the amazingly collaborative and scientifically fruitful, but also extremely fun environment over the years. The most credit for that goes to Alex and after that Daniel, who both welcomed me to the lab from day 1, but also Meryl and Alja and the honorary lab-mate Duncan. I'll forever remember all the trips, parties, concerts and more, but also the hours we spent discussing almost every day, for the afternoon coffees often turning into late night beers spent talking about science, the world and our projects and ideas. I'm also thankful to Caro, Sonja, Tom G. and Tom B., who all helped make and keep the lab great over the years in their own ways.

I'm further thankful to the many other members of the Max Planck Institute of Neurobiology for making it an absolutely amazing place to work and spend time at and my GSN cohort for all the great times and input.

Finally, I'd like to thank my mother Melita, for believing in me and supporting me throughout my academic journey.

8.2 CURRICULUM VITAE

Nejc "Nate" **Dolensek**

ndolensek@neuro.mpg.de | ndolensek.com | NDolensek | ORCID

Education

Ludwig-Maximilians-Universität(LMU), *Graduate School of Systemic Neurosciences(GSN)* — *Munich, Germany*

PhD in Systems Neuroscience — *10/2015-now*

- Selective fast-track programme

University of Groningen & Utrecht University — *The Netherlands*

B.Sc. in Psychology with Honours — *09/2012-08/2015*

- Specialization in neuroscience, minors in mathematics/statistics and affective neuroscience
- Participated in Honours College - Selective **(top 3%)** interdisciplinary, research-focused programme; followed in addition to the B.Sc.

Certifications & More

2017 **Microsoft Professional Program for Data Science**, 12 months, Microsoft *(in progress)*

2015 **Data Science Specialization**, 12 months; Johns Hopkins University, Coursera

2014 **Research Methods in Human Computer Interaction**, PhD Summer School, University of Tallinn — *Tallinn, Estonia*

Publications

Facial expressions of emotion states and their neuronal correlates in mice
Nejc Dolensek, Daniel A. Gehrlach, Alexandra S. Klein, Nadine Gogolla
Science (2020)

Machine-learning approaches to classify and understand emotion states in mice
Nejc Dolensek, Nadine Gogolla
Neuropsychopharmacology (2020)

Aversive state processing in the posterior insular cortex
Daniel A. Gehrlach, **Nejc Dolensek**, Alexandra S. Klein, Ritu Roy Chowdhury, Arthur Matthys, Michaela Junghänel, Thomas N. Gaitanos, Alja Podgornik, Thomas D. Black, Narasimha Reddy Vaka, Karl-Klaus Conzelmann, Nadine Gogolla
Nature Neuroscience (2019)

Research Experience

LMU Munich — *Munich, Germany*

PhD Researcher - PI: Nadine Gogolla, PhD; Circuits for Emotion Group — *11/2016-ongoing*

- Developed and built experimental setups, suite of sensors and analysis pipelines enabling precise stimulus delivery combined with detailed behavioural, physiological and neural observations
- Used computational and machine learning approaches to study emotion processing in the cortex and facial expressions of emotion in mice

Utrecht University — *Utrecht, Netherlands*

Bachelor Thesis Project - *"Data-Driven Analysis of Cognitive Strategies in Ethnic Face Judgements"*, PI: prof. Ron Dotsch — *02/2015-08/2015*

- Designed, programmed and performed psychophysics experiments and used computational modeling to investigate decision making processes behind face classification

University of Groningen — *Groningen, Netherlands*

Bachelor Thesis Project - *"Examining Cognitive Limits of Temporal Integration: A Data-Driven Approach"*, PI: prof. Elkan Akyurek — *09/2014-10/2015*

- Independently designed, programmed and performed a series of psychophysics experiments investigating low-level processes of visual perception

University of Groningen — *Groningen, Netherlands*

Honours Researcher - *"(Un)learning Disgust"*, PI: Charmaine Borg, PhD — *09/2013-09/2014*

- Established protocols to observe disgust conditioning on a subjective, behavioral and physiological(EMG, GSR) level
- Programmed and performed experiments uncovering greatly accelerated conditioned disgust extinction where safety cues are used
- Designed and created an online simulated lab environment to enable a low-cost long-term effects measurement

University of Groningen — *Groningen, Netherlands*

Research Assistant - PI: prof. Jakob Jolij — *01/2013-08/2013*

- Established automated, nonrinvasive, long-distance emotional state monitoring via micro-expression detection

Other Experience

University of Groningen — *Groningen, the Netherlands*

INDEPENDENT TEACHING ASSISTANT & STUDENT MENTOR — *09/2013-04/2015*

- **Courses**: *Data Visualization Workshop, Introduction to Research Methods, Scientific Thinking, Writing and Presenting*

AcademicSolution — *Groningen, the Netherlands*

CONSULTANT & RESEARCHER — *2013*

- Worked on complex, real-life problems from industry in interdisciplinary teams

Kemikalije (Startup) — *Ljubljana, Slovenia*

FOUNDER — *05/2008-01/2013*

- Founded and independently led a successful laboratory equipment and supplies import & e-commerce business
- Cooperated with partners and suppliers from Germany, Poland and China; operated EU-wide
- Employed predictive modelling to manage stock levels and maximise margins

Honors & Awards

ACADEMIC

2020	**Young Scientist Award,**, Max Planck Institute of Neurobiology, awarded to author of best paper in 2020	*~1 000€*
2018	**Best Poster Award**, Gordon Research Conference: Optogenetics and Imaging	*~1 000$*
2015	**1st year stipend**, GSN/LMU Munich, for Fast-track PhD, Awarded to Top Applicants	*~30 000€*
2014	**Stipend**, Tallinn University, to Attend HCI research methods summer school	*~1 000€*
2008-12	**Zois' Scholarship**, Republic of Slovenia, merit scholarship for gifted students (**top 5%**)	*~5 000€*

OTHER

2018	**Best Talk**, Munich Science Slam, *"The Language of Emotion"*	
2017	**Winner & Most Creative Idea**, Burda Hackday, virtual reality with eye tracking for early autism diagnosis	
2017	**Winner & Best Technical Idea**, INNOVATE.HEALTHCARE Hackathon, EEG mind/health-state inference apps	
2012	**Top Award**, Global Blackberry Application Creation Contest, created an activity tracking fitness app	*~2 500€*

Skills

COMPUTER SKILLS

Languages, Python, R, C/C++, Matlab
OS, Windows, Linux
Hardware, NI DAQ, Arduino, Raspberry Pi, Custom circuits, (Wearable) Sensor development
Software, Adobe Photoshop & Illustrator, IBM SPSS, Microsoft Office, Microsoft Visual Studio
Other, Hardware/software interfacing, 3D motion tracking, virtual reality, machine vision, machine learning

LANGUAGES

English, Fluent, C2
German, Basic, A2
Dutch, Basic, A1
Serbian, Croatian, Bosnian, Intermediate, B1
Slovenian, Native

8.3 LIST OF PUBLICATIONS

[*]Dolensek, N., Gehrlach, D. A., Klein, A. S., & Gogolla, N. (2020). Facial expressions of emotion states and their neuronal correlates in mice. *Science, 368*(6486). https://doi.org/10.1126/science.aaz9468

[*]Reprinted with permission from AAAS under AAAS licence to publish (https://www.sciencemag.org/help/reprints-and-permissions)

[^]Gehrlach, D. A., Dolensek, N., Klein, A. S., Roy Chowdhury, R., Matthys, A., Junghänel, M., Gaitanos, T. N., Podgornik, A., Black, T. D., Reddy Vaka, N., Conzelmann, K. K., & Gogolla, N. (2019). Aversive state processing in the posterior insular cortex. *Nature Neuroscience, 22*(9), 1424–1437. https://doi.org/10.1038/s41593-019-0469-1

[^]Reprinted with permission from Springer Nature (Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution; https://www.nature.com/nature-research/reprints-and-permissions/permissions-requests)

[†]Dolensek, N., & Gogolla, N. (2020). Machine-learning approaches to classify and understand emotion states in mice. *Neuropsychopharmacology 2020 46:1*. https://www.nature.com/articles/s41386-020-00857-8

[†]Reprinted with permission from Springer Nature (Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution; https://www.nature.com/nature-research/reprints-and-permissions/permissions-requests)

8.4 DECLARATION OF AUTHOR CONTRIBUTIONS

Facial expressions of emotion states and their neuronal correlates in mice

Nadine Gogolla and Nejc Dolensek conceived the project and designed the experiments. Nejc Dolensek performed all experiments and developed and performed all facial expression analysis. Daniel A. Gehrlach and Alexandra S. Klein helped with the optogenetic experiments. Nadine Gogolla wrote the manuscript with assistance from Nejc Dolensek.

Aversive state processing in the posterior insular cortex

Daniel A. Gehrlach Nadine Gogolla designed the study and analyzed data. Daniel A. Gehrlach, Alexandra S. Klein., Michaela Junghänel and Arthur Matthys performed optogenetic surgeries, behavior experiments and analyses. Nejc Dolensek performed all two-photon imaging experiments and analyses and helped with physiological recordings. Ritu Roy Chowdhury and Alexandra S. Klein. performed fiber photometry recordings. Ritu Roy Chowdhury, Daniel A. Gehrlach and Alexandra S. Klein performed photometry analyses. Alexandra S. Klein. performed optrode recordings. Narasimha Reddy Vaka assisted with behavior analysis and provided custom-written code. Thomas D. Black and Alja Podgornik helped with the histology. Karl-Klaus Conzelmann provided rabies virus and shared expertise in monosynaptic tracing. Daniel A. Gehrlach performed all tracing experiments. Daniel A. Gehrlach and Thomas N. Gaitanos analyzed tracing experiments and performed immunohistochemistry. Nadine Gogolla wrote the manuscript with input from all authors.

Machine-learning approaches to classify and understand emotion states in mice

Nadine Gogolla and Nejc Dolensek wrote this manuscript.

Nate Nejc Dolensek

_____

Nate Nejc Dolensek
Author of this dissertation

Nadine Gogolla

_____

Dr. Nadine Gogolla
Supervisor

8.5 EIDESSTATTLICHE VERSICHERUNG / AFFIDAVIT

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation *Facial expressions of emotion states and their correlates in insular cortex* selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.
I hereby confirm that the dissertation *Facial expressions of emotion states and their correlates in insular cortex* is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

München, den Munich, date	Unterschrift Signature
13.10.2022	*Nate Nejc Dolensek*